ESL-PR339

AD A086070

IMAGE TRANSFORMATION STUDY

R. Devich, F. Weinhaus, L. Martin, D. Collier, G. Kiremidjian,
A. Silvestri, D. McGuirk, B. Gleaton, R. LaPado
ESL, Incorporated
495 Java Drive
Sunnyvale, CA 94086

DTIC
SELECTED
JUL 1 1980
A

30 May 1980

Semi-Annual Technical Report for Period
1 October 1979 - 30 April 1980

Prepared for

80   6 30 080

The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Defense Advanced Research Projects Agency or the U.S. Government.

SEMI-ANNUAL TECHNICAL REPORT NO. 1
IMAGE TRANSFORMATION STUDY

ARPA Order Number:	3854
Name of Contractor:	ESL Incorporated
Effective Date of Contract:	1 November 1979
Contract Expiration Date:	30 November 1980
Contract Number:	MDA903-80-C-0108
Project Manager:	Robert Devich (408) 738-2888 x6656
Short Title of Work:	Image Transformation Study
Contract Period Covered by Report:	1 October 1979 - 30 April 1980

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO. AD-A086070	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Semi-Annual Technical Report No. 1 Image Transformation Study.		5. TYPE OF REPORT & PERIOD COVERED Semi-Annual Technical Rept. 1 Oct 79-Apr 80
		6. PERFORMING ORG. REPORT NUMBER ESL-PR339
7. AUTHOR(s) R. Devich F. Weinhaus L. Martin D. Collier G. Kiremidjian A. Silvestri D. McGuirk, B. Gleaton, R. LaPado		8. CONTRACT OR GRANT NUMBER(s) MDA903-80-C-0108 ARPA Order-3854
9. PERFORMING ORGANIZATION NAME AND ADDRESS ESL, Inc. A Subsidiary of TRW 495 Java Dr., Sunnyvale, CA 94086		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Defense Advanced Research Projects Agency 1400 Wilson Blvd. Arlington, VA 22209		12. REPORT DATE May 30, 1980
		13. NUMBER OF PAGES 72
14. MONITORING AGENCY NAME & ADDRESS(If different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

ESL-PR339

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Digital Image Processing
Image Transformations
Perspective Transformations
Three Dimensional Graphics

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

See Report Summary, Section 1

404778

DD FORM 1473 EDITION OF 1 NOV 65 IS OBSOLETE
1 JAN 73

TABLE OF CONTENTS

iii

LIST OF ILLUSTRATIONS

LIST OF TABLES

1.   REPORT SUMMARY.

The purpose of the Image Transformation Study is to develop techniques for converting oblique aerial photography to the ground-level viewing perspectives that would be seen by an observer as he moves around within the imaged area. This report documents the first six months of the investigation.

The approach to the problem included five areas:

- A literature search was performed to identify previous investigators' progress.

- Appropriate aerial imagery was acquired and digitized.

- The mathematical relationships defining the required image transformations were derived.

- Computer programs were developed to perform the mathematical manipulations.

- Representative ground-level perspective views were created using the computer programs.

Aerial photographs were taken over three areas of Santa Clara County. Photographs of each area were taken from four oblique directions as well as overhead, and were taken in both morning and afternoon light to ensure optimum viewing conditions. The 12-inch focal length camera was at 3000 to 3500 foot altitudes and used 9-inch wide black and white film. Portions of the film have been digitized (i.e., converted to the format of a digital image processing computer system) at a resolution corresponding to a 3-inch ground picture element size. The digitized imagery has been used in subsequent stages of the study. Section 6.1 has examples of the aerial photography.

1.   -- Continued.

A mathematical technique for deriving a camera model (i.e., the camera's position and orientation) from a photograph taken by the camera in conjunction with ground truth data was developed and applied in order to obtain this information used in the image transformations. Section 3 presents the technique.

The geometric projections used in this study are perspective transformations. The one-camera perspective transformations, which relate image and spatial information, were derived. The two-camera perspective transformations, which relate imagery viewed from one camera position to that which would be viewed from a second camera position, were also derived. The perspective transformations are presented in Section 2.

A computer program was written to interactively extract and manipulate a stick figure model that represents the underlying three-dimensional structure of a photographed scene. The three dimensional stick figure model contains the information needed to convert digital imagery from one perspective view to another. The stick figure model program is described in Section 4.

A second computer program was written to perform the perspective transformations of the gray level or intensity information in a digital image to convert it from the overhead oblique viewing position to ground level horizontal viewing positions. This program is discussed in Section 5.

The computer programs were used on the digitized images to produce several examples of output images. These images are realistic in their rendition of the scenes and maintain the same resolution as the original imagery. The examples are presented in Section 6.

2. PERSPECTIVE EQUATIONS.

The perspective equations are the mathematical description of the process of imaging with a system of ideal cameras. An ideal camera has a flat focal surface, called the imaging plane. A point in space (the object point) is imaged onto this plane at the intersection of the line from the lens point to the object point with the imaging plane (the image point).

The one-camera perspective equations (Section 2.1) relate the object point to the image point. The two-camera perspective equations (Section 2.2) relate the image point in one ideal camera to the image point in a second ideal camera.

2.1 One Camera Perspective Equations.

2.1.1 Definitions.

The perspective equations describe the photographic process; namely, the conversion of three dimensional data into a two dimensional representation. Figure 2-1 shows a point (x,y,z) in the spatial or global coordinate system and the corresponding picture coordinates (x_p, z_p) as imaged by the camera. Note that the picture coordinate system is centered on the optical axis a distance equal to the focal length from the lens of the camera. The camera is positioned at spatial coordinates (x_o, y_o, z_o) and has an orientation given by its pan, tilt and roll angles (θ, ϕ, ζ). These angles are defined as follows. Consider a coordinate system with the camera at the origin and pointing along the Y axis. Pan is a rotation of the coordinate system about its Z axis. Tilt is a rotation of the coordinate system about its X axis. Roll is the rotation of the coordinate system about its Y axis. The rotations are counter clockwise and must be done in this order. The overall transformation of the camera from its initial orientation to its present orientation can be expressed as a rotation matrix, namely
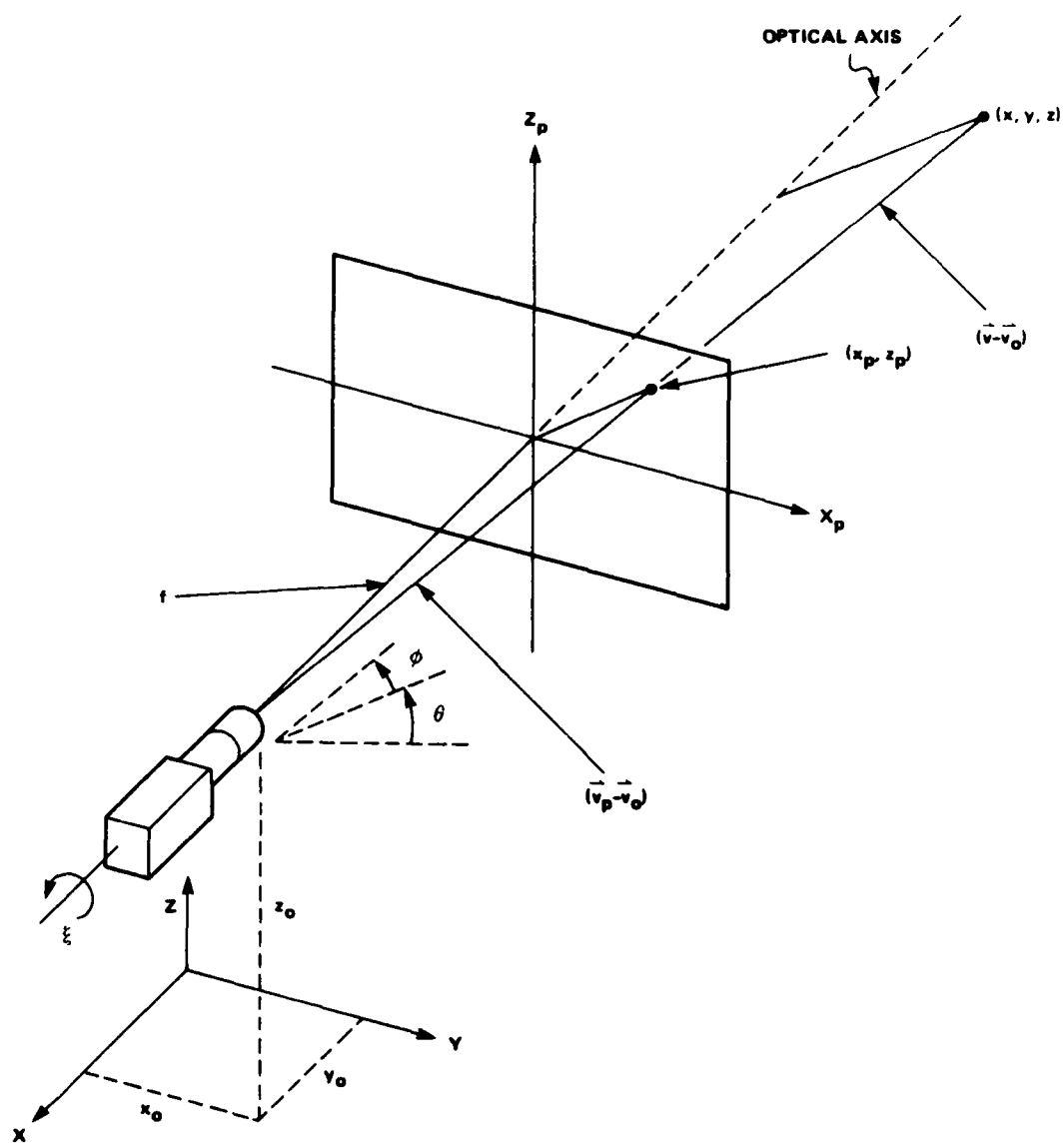
Figure 2-1. Projection Geometry

2.1.1 -- Continued.

$$
\mathbf{R} = \zeta\theta\phi = \begin{bmatrix} \cos\zeta & 0 & \sin\zeta \\ 0 & 1 & 0 \\ -\sin\zeta & 0 & \cos\zeta \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & \sin\phi \\ 0 & -\sin\phi & \cos\phi \end{bmatrix} \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}
$$

$$
= \begin{bmatrix} \cos\zeta\,\cos\theta + \sin\zeta\,\sin\phi\,\sin\theta & \cos\zeta\,\sin\theta - \sin\zeta\,\sin\phi\,\cos\theta & \sin\zeta\,\cos\phi \\ -\cos\phi\,\sin\theta & \cos\phi\,\cos\theta & \sin\phi \\ -\sin\zeta\,\cos\theta + \cos\zeta\,\sin\phi\,\sin\theta & -\sin\zeta\,\sin\theta - \cos\zeta\,\sin\phi\,\cos\theta & \cos\zeta\,\cos\phi \end{bmatrix}
$$

$$(2-1)$$

$\mathbf{R}$ is a unitary matrix, i.e., $\mathbf{R}^{-1} = \mathbf{R}^T$. The columns of $\mathbf{R}^{-1}$ represent the unit vectors in the camera coordinate system (the picture coordinate system) relative to the spatial coordinate system:

$$
\hat{e}_1 \equiv \begin{bmatrix} \cos\zeta\,\cos\theta + \sin\zeta\,\sin\phi\,\sin\theta \\ \cos\zeta\,\sin\theta - \sin\zeta\,\sin\phi\,\cos\theta \\ \sin\zeta\,\cos\phi \end{bmatrix}, \tag{2-2}
$$

$$
\hat{e}_2 \equiv \begin{bmatrix} -\cos\phi\,\sin\theta \\ \cos\phi\,\cos\theta \\ \sin\phi \end{bmatrix}, \tag{2-3}
$$

$$
\hat{e}_3 \equiv \begin{bmatrix} -\sin\zeta\,\cos\theta + \cos\zeta\,\sin\phi\,\sin\theta \\ -\sin\zeta\,\sin\theta - \cos\zeta\,\sin\phi\,\cos\theta \\ \cos\zeta\,\cos\phi \end{bmatrix} \tag{2-4}
$$

2.1.2 Forward Transformation.

Any spatial point (x,y,z) can be described by a vector from the spatial origin to the point in question. Thus

2.1.2   -- Continued.

$$\vec{v} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \tag{2-5}$$

and the camera position can be described by

$$\vec{v}_o = \begin{bmatrix} x_o \\ y_o \\ z_o \end{bmatrix} \tag{2-6}$$

If the point lies in the image plane, then it can be expressed as

$$\vec{v}_p = \vec{v}_o + x_p \hat{e}_1 + f \hat{e}_2 + z_p \hat{e}_3 \tag{2-7}$$

where x_p and z_p are the coordinates of the point in the image plane. This equation simply says that the vector $\vec{v}_p$ can be expressed as the vector sum of $\vec{v}_o$ plus the offset of the point from the camera position.

Now consider the triangles formed by the optical axis, the ray from the camera through point (x_p, o, z_p) to point (x,y,z) and the perpendiculars dropped from these points to the optical axis, (see Figure 2-1). These two triangles are similar. Thus

$$\frac{|\vec{v}_p - \vec{v}_o|}{f} = \frac{|\text{projection of } (\vec{v}-\vec{v}_o) \text{ perpendicular to the optical axis}|}{|\text{projection of } (\vec{v}-\vec{v}_o) \text{ along the optical axis}|}$$

$$\tag{2-8}$$

2-4

2.1.2   -- Continued.

In a similar fashion, the components of the vectors $(\vec{v}_p - \vec{v}_o)$ and $(\vec{v} - \vec{v}_o)$ projected relative to the picture coordinate system also form similar triangles. Thus

$$x_p = \frac{f \langle \hat{e}_1, \vec{v} - \vec{v}_o \rangle}{\langle \hat{e}_2, \vec{v} - \vec{v}_o \rangle} \qquad (2-9)$$

and

$$z_p = \frac{f \langle \hat{e}_3, \vec{v} - \vec{v}_o \rangle}{\langle \hat{e}_2, \vec{v} - \vec{v}_o \rangle} \qquad (2-10)$$

where the bracket notation represents the inner product. Equations (2-9) and (2-10) are the single-camera perspective equations that describe the transformation from the spatial coordinate system to the picture coordinate system.

These equations can also be recast into a different representation. Consider the matrix **N** formed by concatenating the transposed unit vectors, namely

$$\mathbf{N} \equiv \begin{bmatrix} \hat{e}_1^{\,T} \\ \hat{e}_3^{\,T} \\ \hat{e}_2^{\,T} \end{bmatrix} \qquad \text{or}$$

$$\left[N_{ij} \right] = \begin{bmatrix} \cos\zeta \cos\theta + \sin\zeta \sin\phi \sin\theta & \cos\zeta \sin\theta - \sin\zeta \sin\phi \cos\theta & \sin\zeta \cos\phi \\ -\sin\zeta \cos\theta + \cos\zeta \sin\phi \sin\theta & -\sin\zeta \sin\theta - \cos\zeta \sin\phi \cos\theta & \cos\zeta \cos\phi \\ -\cos\phi \sin\theta & \cos\phi \cos\theta & \sin\phi \end{bmatrix}$$

2.1.2   -- Continued.

$$\equiv \begin{bmatrix} N_{11} & N_{12} & N_{13} \\ N_{21} & N_{22} & N_{23} \\ N_{31} & N_{32} & N_{33} \end{bmatrix} \tag{2-11}$$

Equations (2-9) and (2-10) can now be expressed as

$$x_p = \frac{f \sum_j N_{1j} \ (\vec{v}-\vec{v}_o)_j}{\sum_j N_{2j} \ (\vec{v}-\vec{v}_0)_j}$$

$$= \frac{f\left[N_{11}(x-x_o) + N_{12}(y-y_o) + N_{13}(z-z_o)\right]}{\left[N_{21}(x-x_o) + N_{22}(y-y_o) + N_{23}(z-z_o)\right]} \tag{2-12}$$

$$z_p = \frac{f \sum_j N_{3j} \ (\vec{v}-\vec{v}_o)_j}{\sum_j N_{2j} \ (\vec{v}-\vec{v}_o)_j}$$

$$= \frac{f\left[N_{31}(x-x_o) + N_{32}(y-y_o) + N_{33}(z-z_o)\right]}{\left[N_{21}(x-x_o) + N_{22}(y-y_o) + N_{23}(z-z_o)\right]} \tag{2-13}$$

2.1.3   Inverse Transformation.

Given a point (x_p, z_p) in the picture coordinate system, it is now desired to find the corresponding point $\vec{v}$ in spatial coordinates. Equations (2-9) and (2-10) can be expressed as

$$< (x_p \hat{e}_2 - f\hat{e}_1), \ (\vec{v}-\vec{v}_o)> = 0 \tag{2-14}$$

2.1.3   -- Continued.

and

$$< (z_p \hat{e}_2 - f\hat{e}_3), \ (\vec{v}-\vec{v}_o) > \ = \ 0 \qquad\qquad (2-15)$$

From these equations it is apparent that the vector $(\vec{v}-\vec{v}_o)$ is orthogonal to both $(x_p\hat{e}_2 - f\hat{e}_1)$ and $(z_p\hat{e}_2 - f\hat{e}_3)$, and is therefore proportional to their cross product:

$$\vec{v}-\vec{v}_o = \mu(x_p\hat{e}_2 - f\hat{e}_1) \ \textbf{X} \ (z_p\hat{e}_2 - f\hat{e}_3) \ , \qquad\qquad (2-16)$$

where μ is an arbitrary constant of proportionality. Using the orthonormality property of the unit vectors, one then obtains

$$\vec{v} = \vec{v}_o + \lambda(x_p\hat{e}_1 + z_p\hat{e}_3 + f\hat{e}_2) \ , \qquad\qquad (2-17)$$

which is the inverse transformation required. The free parameter $\lambda = -\mu f$ represents a distance function along the optical axis and occurs because a 2-D vector is being projected into 3-D space.

    Equation (2-17) can be expressed in a different form for use in subsequent sections of this report. Define

$$\tilde{v}_p \ \equiv \ \begin{bmatrix} x_p \\ z_p \\ f \end{bmatrix}.$$

The tilde is used to distinguish this vector from the spatial vector described by equation (2-7). It represents the picture coordinates of the point relative to the camera. If we now define a matrix

$$\textbf{M} = \textbf{N}^{-1} = \textbf{N}^T, \ \text{i.e.}$$

2.1.3   -- Continued.

$$\left[M_{ij}\right] = \begin{bmatrix} M_{11} & M_{12} & M_{13} \\ M_{21} & M_{22} & M_{23} \\ M_{31} & M_{32} & M_{33} \end{bmatrix} = \begin{bmatrix} N_{11} & N_{21} & N_{31} \\ N_{12} & N_{22} & N_{32} \\ N_{13} & N_{23} & N_{33} \end{bmatrix} , \qquad (2\text{-}18)$$

Then equation (2-17) can be expressed as

$$\vec{v} = \vec{v}_o + \lambda \mathbf{M} \tilde{v}_p \qquad (2\text{-}19)$$

or

$$x = x_o + \lambda \sum_j M_{1j} (v_{p_j})$$
$$= x_o + \lambda (M_{11} x_p + M_{12} z_p + M_{13} f) \qquad (2\text{-}20)$$

$$y = y_o + \lambda \sum_j M_{2j} (v_{p_j})$$
$$= y_o + \lambda (M_{21} x_p + M_{22} z_p + M_{23} f) \qquad (2\text{-}21)$$

$$z = z_o + \lambda \sum_j M_{3j} (v_{p_j})$$
$$= z_o + \lambda (M_{31} x_p + M_{32} z_p + M_{33} f) \qquad (2\text{-}22)$$

2.2   Two-Camera Perspective Equations.

The central concept explored in the study is the transformation of imagery taken from one position to what would be seen from a second position. The mathematical relationships expressing this transformation are the two camera perspective equations.

2.2 _____ -- Continued.

Given a camera at position $\vec{v}_o$, with a focal length f, forward projection matrix **N**, and inverse projection matrix **M**, and a second camera at position $\vec{v}_o'$, focal length f´, and projection matrices **N´** and **M´**, it is desired to derive an expression for the position $\tilde{v}_p'$ in the second camera's focal plane corresponding to an object imaged at $\tilde{v}_p$ in the first camera's focal plane. It is impossible to solve this problem without making assumptions about the scene being imaged, because there are infinitely many points in space which may be imaged at a particular point $\tilde{v}_p$, generating an infinite locus of points $\tilde{v}'_p$. Figure 2-2 (a) shows that $\tilde{v}_p$ corresponds to a ray in space; the ray generates a line in the second camera's focal plane. If we assume that the scene contains a known surface shape, then the problem becomes solvable, as a unique spatial point is imaged at $\tilde{v}_p$; if there are no obstructions, the same point is imaged at $\tilde{v}_p'$. The situation where the imaged scene is a plane is depicted in Figure 2-2 (b); it can be seen that the problem is now solvable. Making the planar assumption, the following derivation applies.

Equation (2-19) states that any point v on the ray imaged at $\tilde{v}_p$ can be represented by

$$\vec{v} = \vec{v}_o + \lambda \mathbf{M}\tilde{v}_p \tag{2-19}$$

The plane can be represented by

$$Ax + By + Cz + D = 0 \tag{2-23}$$

where A, B, C, and D are real numbers, or in vector notation as

$$\langle \vec{a}, \vec{v} \rangle + D = 0 \tag{2-24}$$
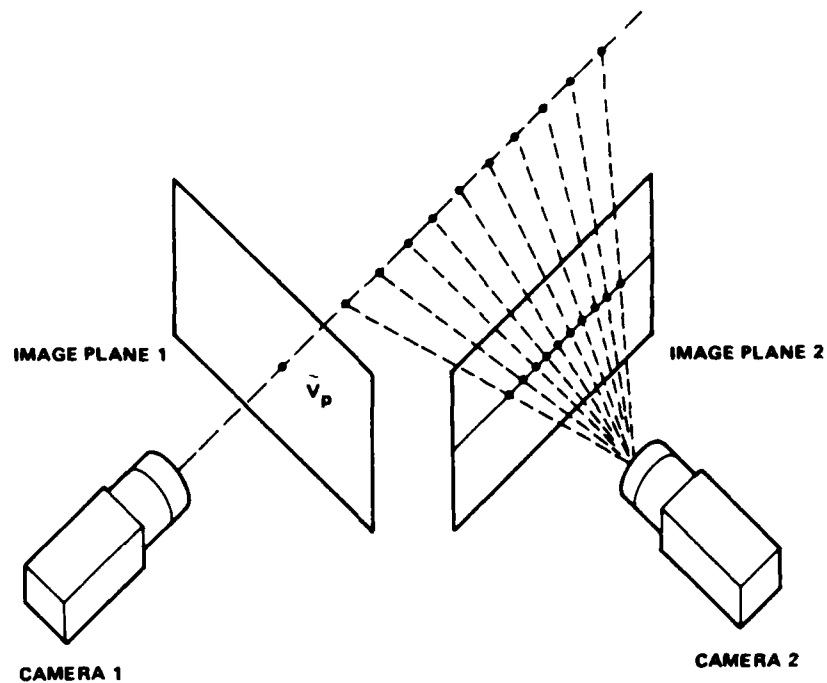
a. The ray defined by $\tilde{v}_p$ generates a line in the second-camera focal plane.

b. Imaging a plane, $\tilde{v}_p$ uniquely determines $\tilde{v}_p'$

Figure 2-2. Two-Camera Perspective Imaging

2.2 -- Continued.

where $\vec{a} \equiv (A,B,C,)^T$, and $\vec{v} = (x,y,z)^T$ is the point at which the imaging ray intersects the plane.

Substituting (2-19) in (2-24) and solving,

$$\lambda = -\frac{<\vec{a},\vec{v}_o>+D}{<\vec{a},M\,\tilde{v}_p>} \ . \tag{2-25}$$

Substituting (2-25) into (2-19),

$$v = v_o - \frac{<\vec{a},\vec{v}_o>+D}{<\vec{a},M\,\tilde{v}_p>}\ M\,\tilde{v}_p \ . \tag{2-26}$$

The second camera images $\vec{v}$ at a point $\tilde{v}'_p$ in its focal plane:

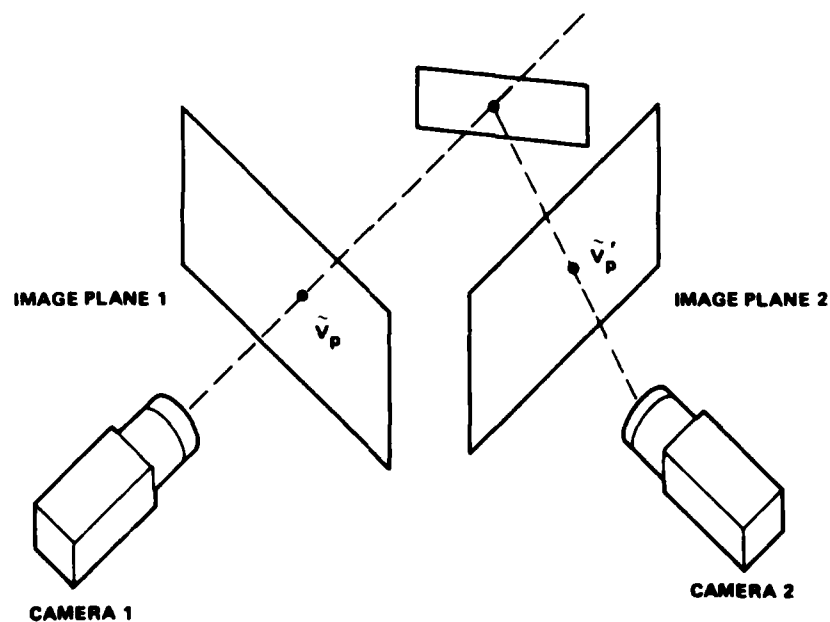$$\tilde{v}'_p = \tilde{v}_o - \frac{<\vec{a},\vec{v}'_o>+D}{<\vec{a},M'\tilde{v}'_p>}\ M'\tilde{v}'_p \tag{2-27}$$

Equating (2-26) and (2-27) and solving,

$$\tilde{v}'_p = \frac{<\vec{a},M\,\tilde{v}'_p>}{<\vec{a},\vec{v}_o>+D}\ M'^{-1}\left[(\vec{v}_o-\vec{v}'_o) - \frac{<\vec{a},\vec{v}_o>+D}{<\vec{a},M\,\tilde{v}_p>}\ M\,\tilde{v}_p\right]. \tag{2-28}$$

Since the third coordinate of $\tilde{v}'_p$ is f', equation (2-28) implies that

$$f' = \frac{<\vec{a},M\,\tilde{v}'_p>}{<\vec{a},\vec{v}'_o>+D}\ M_3'^{-1}, \ \vec{v}_o-\vec{v}'_o - \frac{<\vec{a},\vec{v}_o>+D}{<\vec{a},M\,\tilde{v}_p>}\ M\,\tilde{v}_p \ , \tag{2-29}$$

2.2 -- Continued.

where $M_3^{\prime -1}$ represents the row vector made up of the elements of the third row of the matrix $\mathbf{M}^{\prime -1}$. From (2-29),

$$\frac{<\vec{a}, \mathbf{M}\tilde{v}_p>}{<\vec{a},\vec{v}_o^{\prime}>+D} = \frac{f^{\prime}}{<M_3^{\prime -1}, \vec{v}_o-\vec{v}_o^{\prime} - \dfrac{<\vec{a},\vec{v}_o>+D}{<\vec{a},\mathbf{M}\tilde{v}_p>} \mathbf{M}\tilde{v}_p>} \quad . \tag{2-30}$$

Substituting (2-30) into (2-28),

$$\tilde{v}_p^{\prime} = \frac{f^{\prime} \mathbf{M}^{\prime -1}(<\vec{a}, \mathbf{M}\tilde{v}_p>(\vec{v}_o-\vec{v}_o^{\prime}) - (<\vec{a},\vec{v}_o>+D)\mathbf{M}\tilde{v}_p)}{<M_3^{\prime -1},<\vec{a}, \mathbf{M}\tilde{v}_p>(\vec{v}_o-\vec{v}_o^{\prime}) - (<\vec{a},\vec{v}_o>+D)\mathbf{M}\tilde{v}_p>} \quad , \tag{2-31}$$

which can also be written as

$$x_p^{\prime} = \frac{f^{\prime}<M_1^{\prime -1},<\vec{a}, \mathbf{M}\tilde{v}_p>(\vec{v}_o-\vec{v}_o^{\prime}) - (<\vec{a},\vec{v}_o>+D)\mathbf{M}\tilde{v}_p>}{<M_3^{\prime -1},<\vec{a}, \mathbf{M}\tilde{v}_p>(\vec{v}_o-\vec{v}_o^{\prime}) - (<\vec{a},\vec{v}_o>+D)\mathbf{M}\tilde{v}_p>} \tag{2-32}$$

$$z_p^{\prime} = \frac{f^{\prime}<M_2^{\prime -1},<\vec{a}, \mathbf{M}\tilde{v}_p>(\vec{v}_o-\vec{v}_o^{\prime}) - (<\vec{a},\vec{v}_o>+D)\mathbf{M}\tilde{v}_p>}{<M_3^{\prime -1},<\vec{a}, \mathbf{M}\tilde{v}_p>(\vec{v}_o-\vec{v}_o^{\prime}) - (<\vec{a},\vec{v}_o>+D)\mathbf{M}\tilde{v}_p>} \tag{2-33}$$

Equations (2-32) and 2-33) are the two camera perspective equations. They are fractional linear in form (i.e., the numerator and denominator are linear functions of x_p and z_p). The equations are valid as long as the plane does not contain either camera.

3.   IMAGE MODEL.

Black and white aerial photography was taken with a 12-inch focal length RC-10 camera at 3000 to 3500 feet altitude and depression angles of 50 to 60 degrees from horizontal. The imagery was digitized with 10-bit precision and 10 to 30 μm pixel size with a PDS-1010 scanning microdensitometer.

The image model is the ancillary information used to fully define the relationship of a digital image to a spatial coordinate system. This information consists of the camera model (the position, orientation, and focal length of the camera) and the scanning parameters (the position of the microdensitometer origin with respect to the image center and the pixel spacing). This section discusses the acquisition of the camera model and the scanning parameters.

3.1   Camera Model Derivation and Extraction.

The problem of identifying the position and orientation of a camera from its photography is known as resectioning. A number of techniques for resectioning are described in the literature of photogrammetry.

General solutions have been found that require iterative techniques.[1,2,3] They require an initial approximation to the three spatial coordinates and three orientation angles of the camera. At least three pairs of conjugate picture and spatial control points are required and used with the perspective equations in a least squares procedure to find corrections to the initial approximations.

Closed form solutions have also been developed.[4,5] However, they make the rather limiting assumption of nearly downlooking views.

3.1 _____ -- Continued.

We have developed an approach that is straight-forward and
simple in its application, but requires a rectangular parallelepiped
structure in the photograph located near the optical axis. It
requires that one pair of conjugate control points be measured in
the image and on the ground, in addition to the coordinate system
origin.

Consider Figure 3-1, which shows an image of a rectangular
parallelepiped structure located on horizontal terrain. When projected
into the image, the spatial corner point labelled p_o, is located at the
optical center of the photograph which is the origin of the picture
coordinate system. The origin of the spatial coordinate system is
centered at the corner of the structure, with its axes parallel to the
sides as shown. If the angles Ω and Ψ and the slope of the vertical
edge defined by $\overline{p_o p_3}$ are measured in the photograph, then knowledge of
these parameters along with the perspective equations applied to the
spatial points p_1, p_2 and p_3 is sufficient to solve for the pan, tilt,
and roll angles θ, ϕ, ζ describing the camera orientation as follows.

Consider the projection of the spatial points p_1, p_2 and p_3
into the image. From the perspective equations (2-9) and (2-10),
these image points can be expressed as $p_{pi} = (x_{pi}, z_{pi})$, where

$$
\left.
\begin{aligned}
x_{pi} &= C_i \langle \hat{e}_1, \vec{v}_i - \vec{v}_o \rangle \\[2ex]
z_{pi} &= C_i \langle \hat{e}_3, \vec{v}_i - \vec{v}_o \rangle
\end{aligned}
\right\}
\qquad i = 1,2,3
\qquad\qquad (3\text{-}1)
$$

In these expressions $C_i = f / \langle \hat{e}_2, \vec{v}_i - \vec{v}_o \rangle$, and $\vec{v}_i$ is the vector from the
origin to the spatial points p_i. Since the spatial origin is at p_o,
the vector from the origin to the camera position, $\vec{v}_o$, will be parallel
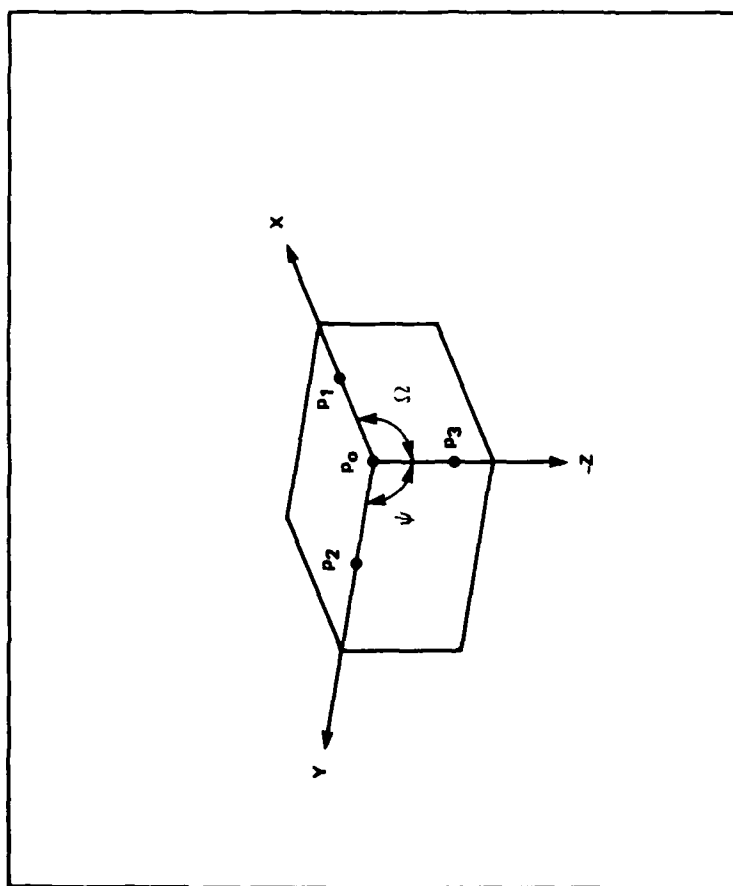to the unit vector $\hat{e}_2$. Thus

Figure 3-1. Orientation Angle Geometry

3.1   -- Continued.

$$\langle \hat{e}_1, \vec{v}_o \rangle = \langle \hat{e}_3, \vec{v}_o \rangle = 0 \qquad (3-2)$$

If $\vec{p}_{pi}$ are defined to be vectors from the picture origin to the picture points p_{pi}, then using equations (3-1) and (3-2) one obtains

$$\vec{p}_{pi} = \begin{bmatrix} x_{pi} \\ z_{pi} \end{bmatrix} = C_i \begin{bmatrix} \langle \hat{e}_1, \vec{v}_i \rangle \\ \langle \hat{e}_3, \vec{v}_i \rangle \end{bmatrix} \qquad ; \qquad i = 1,2,3 \quad . \qquad (3-3)$$

Choosing $\vec{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ $\vec{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ and $\vec{v}_3 = \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix}$

leads to

$$\vec{p}_{pi} = C_i \begin{bmatrix} \cos\zeta \cos\theta + \sin\zeta \sin\phi \sin\theta \\ -\sin\zeta \cos\theta + \cos\zeta \sin\phi \sin\theta \end{bmatrix} \qquad , \qquad (3-4)$$

$$\vec{p}_{p2} = C_i \begin{bmatrix} \cos\zeta \sin\theta - \sin\zeta \sin\phi \cos\theta \\ -\sin\zeta \sin\theta - \cos\zeta \sin\phi \cos\theta \end{bmatrix} \qquad , \qquad (3-5)$$

and

$$\vec{p}_{p3} = -C_i \begin{bmatrix} \sin\zeta \cos\phi \\ \cos\zeta \cos\phi \end{bmatrix} \quad . \qquad (3-6)$$

From equation (3-6)

$$\tan^{-1} \left(\frac{x_{p3}}{z_{p3}} \right) = \tan^{-1} \left(\frac{\sin\zeta}{\cos\zeta} \right) = \zeta \qquad (3-7)$$

3.1   -- Continued.

This expression indicates that the slope of the vertical edge of the building $\overline{P_0P_3}$, as measured in the image, identifies the roll angle, ζ, of the camera.

The remaining orientation angles can be obtained from the angles Ω and Ψ measured in the photograph. These angles can be expressed from the definition of the inner product as

$$\cos\Omega = \frac{<\vec{P}_{p1} , \vec{P}_{p3}>}{|\vec{P}_{p1}||\vec{P}_{p3}|} \qquad (3-8)$$

and

$$\cos\Psi = \frac{<\vec{P}_{p2} , \vec{P}_{p3}>}{|\vec{P}_{p2}||\vec{P}_{p3}|} \qquad (3-9)$$

Substituting equations (3-4), (3-5), and (3-6) into equations (3-8) and (3-9), one obtains

$$\cos\Omega = \frac{-\sin\phi \sin\theta}{(\cos^2\theta + \sin^2\phi \sin^2\theta)^{1/2}} \equiv \kappa_1 \qquad (3-10)$$

and

$$\cos\Psi = \frac{\sin\phi \cos\theta}{(\sin^2\theta + \sin^2\phi \cos^2\theta)^{1/2}} \equiv \kappa_2 \quad . \qquad (3-11)$$

Squaring both sides of equations (3-10) and (3-11) and solving for the numerator then leads to

3.1   -- Continued.

$$\sin^2\phi \; \sin^2\theta = \kappa_1^2 \cos^2\theta + \kappa_1^2 \sin^2\phi \; \sin^2\theta \tag{3-12}$$

and

$$\sin^2\phi \; \cos^2\theta = \kappa_2^2 \sin^2\theta + \kappa_2^2 \sin^2\phi \; \cos^2\theta \quad . \tag{3-13}$$

From equation (3-12) one obtains

$$\sin^2\phi = \frac{\kappa_1^2 \cot^2\theta}{1-\kappa_1^2} \quad . \tag{3-14}$$

Similarly from equation (3-13) one obtains

$$\sin^2\phi = \frac{\kappa_1^2 \tan^2\theta}{1-\kappa_2^2} \quad . \tag{3-15}$$

Equating (3-14) and (3-15) and solving for θ,

$$\tan^4\theta = \frac{\kappa_1^2}{\kappa_2^2} \; \frac{1-\kappa_2^2}{1-\kappa_1^2} = \frac{\cos^2\Omega \; \sin^2\Psi}{\cos^2\Psi \; \sin^2\Omega} \tag{3-16}$$

or

$$\theta = - \tan^{-1} \left[\frac{\tan\Psi}{\tan\Omega}\right]^{1/2} \quad , \tag{3-17}$$

3-6

3.1   -- Continued.

where the negative root is chosen to be consistent with the definition
of the pan angle.

    If equation (3-16) is substituted into the square of equation
(3-15) one obtains

$$\sin^4\phi = \frac{1}{\tan^2\Omega \ \tan^2\Psi} \qquad (3\text{-}18)$$

or

$$\phi = -\sin^{-1}\left[\frac{1}{\tan\Omega \ \tan\Psi}\right]^{1/2} \qquad (3\text{-}19)$$

Where the negative root is chosen to be consistent with the definition
of the tilt angle. (Note that $\tan\Omega$ and $\tan\Psi$ are both negative, since
for the orientation of the building shown, both angles are greater
than 90°.)

    The above formalism was developed for the special case in
which the optical center was coincident with the projection of the
closest corner of the building as shown. This formalism can be
extended to any corner of the building for which the three inter-
secting edges of the building are visible. If p_1, p_2, and p_3 are
always chosen as $p_1 = (\pm1,0,0)$ $p_2 = (0,\pm1,0)$ and $p_3 = (0,0,\pm1)$ then
equations (3-5), (3-17), and (3-19) become

$$\zeta = \tan^{-1}\left(\frac{x_{p3}}{z_{p3}}\right) \qquad (3\text{-}20)$$

3.1   -- Continued.

$$\theta = \pm\tan^{-1}\left[\left|\frac{\tan\Psi}{\tan\Omega}\right|\right]^{1/2} \tag{3-21}$$

and

$$\phi = -\sin^{-1}\left[\frac{1}{|\tan\Omega| \, |\tan\Psi|}\right]^{1/2} \tag{3-22}$$

The negative sign for equation (3-22) is always correct, since the
input scenes are always downlooking and thus have a negative tilt.
The sign for equation (3-21) must be determined by the orientation
of the projection of the Y spatial axis in the photograph. If it
points towards the left side of the photograph, θ will be negative.
If it points towards the right side of the photograph, then θ will
be positive.

Once the camera orientation angles have been identified,
the camera position can be obtained from an application of the
inverse perspective equations with at least one pair of conjugate
picture and spatial coordinates. The picture coordinates can be
measured either directly from the original photograph, or in the
digital image and converted to picture coordinates as will be
described in the following section. The spatial coordinates are
measured on a map and are generally at ground level.

If one takes $\vec{v} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$, the spatial origin, then $\tilde{v}_p = \begin{bmatrix} 0 \\ 0 \\ f \end{bmatrix}$,

since the projection of the spatial origin lies at the optical center.
Substitution of these values into the inverse perspective equation,
equation (2-17), leads to

$$\vec{v}_o = -\lambda f\hat{e}_2, \tag{3-23}$$

3.1   -- Continued.

which indeed shows that $\vec{v}_o$ is parallel to $\hat{e}_2$. When equation (3-23) is substituted into the perspective equations, equations (2-9) and (2-10), one obtains

$$\lambda = \frac{f\langle \hat{e}_1, \vec{v}\rangle - x_p\langle \hat{e}_2, \vec{v}\rangle}{fx_p} \tag{3-24}$$

and

$$\lambda = \frac{f\langle \hat{e}_3, \vec{v}\rangle - z_p\langle \hat{e}_2, \vec{v}\rangle}{fz_p} . \tag{3-25}$$

The parameter λ can be obtained by substitution of at least one pair of conjugate spatial coordinates, $\vec{v} = (x,y,z)^T$, and picture coordinates, $\vec{v}_p = (x_p,z_p)$, into equations (3-24) and (3-25) and averaging the results. The camera location, $\vec{v}_o = (x_o,y_o,z_o)^T$, then can be determined by substitution of this value for λ into equation (3-23).

If the spatial origin is selected at ground level, then the analysis of the camera position follows by a straight-forward application of the above, namely equations (3-23), (3-24), and (3-25). If however, the spatial origin is above ground level, as shown in Figure 3-1, then the coordinates of the spatial control points will be at values below $z = 0$. This non-zero offset is equal to the height of the building at the spatial origin. In most cases, this height is not available on the map. It can be determined in at least two ways. The first way is to perform shadow mensuration, provided that a clearly visible shadow is available for the building in question. The second method is an iterative approach. First, assume the origin is at ground level. Obtain the camera position as described above.

3.1 -- Continued.

Then use the 3-D stick figure model algorithm (to be described below)
with this camera model to find the height of the building. Iterate this
technique to the desired accuracy.

3.2 Scanning Parameters.

In addition to the camera parameters, certain scanning
parameters are needed to relate the digital image coordinate system
measured in lines and samples into picture coordinates required by
the perspective equations. These parameters are: the sample spacing
and the offsets between the optical center and the upper left corner
of the digital image subsection scanned.

Figure 3-2 depicts a piece of film, a small section of which
is to be scanned with a microdensitometer and converted into digital
form. There are three basic coordinate systems to be considered, each
of which is shown in this figure.

The first coordinate system is that of the microdensitometer
whose coordinate axes are labelled X_s and Z_s. Note that the direction
of the X_s axis is opposite to the standard format. The units of this
system are in micrometers and the origin is arbitrary.

The second coordinate system is the picture coordinate system
whose axes are labelled X_p and Z_p. It also is measured in units of
micrometers and its origin is at the optical center of the image. The
perspective equations are described in terms of this coordinate system.

The third coordinate system is that of the digital image.
It is measured in pixels along the line (L) and sample (S) directions.
Pixel (1,1) is the upper left hand corner of the digital image.
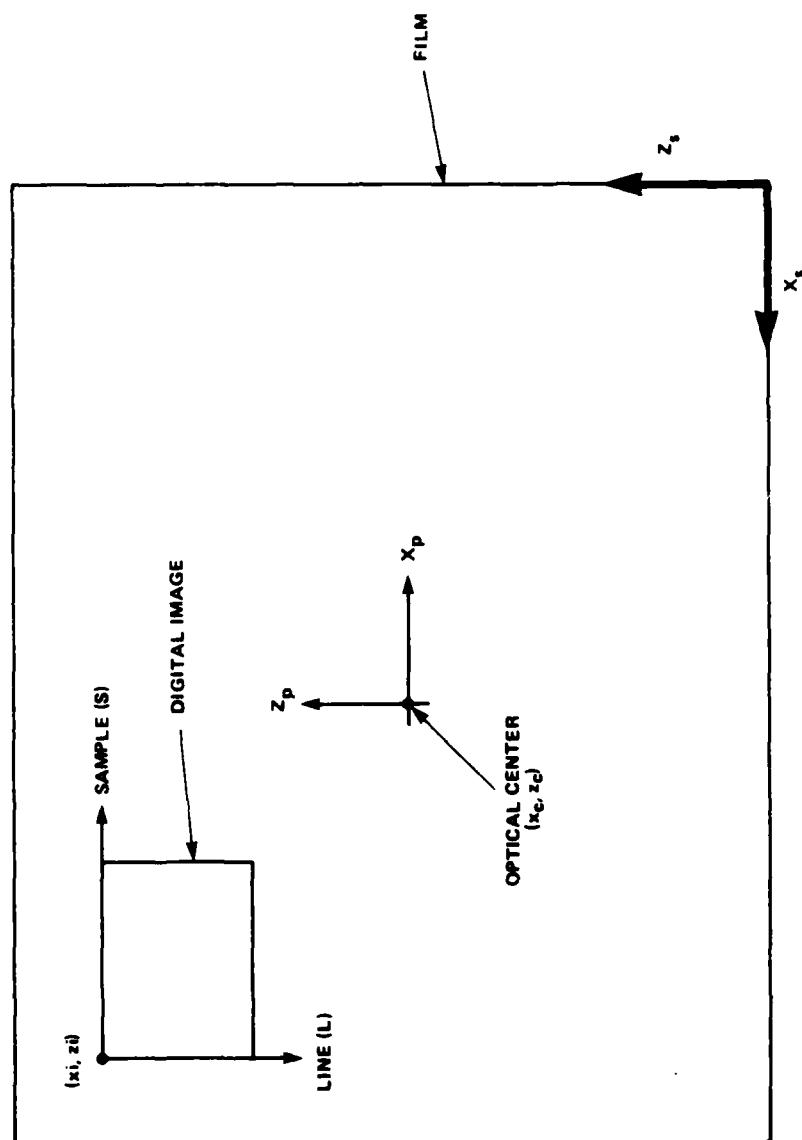
Figure 3-2. Scanning Coordinate System

3.2   -- Continued.

When a picture is scanned by a microdensitometer, a light spot of fixed size is moved across the film at fixed intervals or steps. For each placement of the spot, the light intensity transmitted (or reflected) is coded as the intensity or gray level of the image at that position. The step size or interval moved between measurements is called the sample spacing (SS). The number of measurements in the horizontal directions indicates the sample dimension in pixels and the number of measurements in the vertical direction identifies the line dimension in pixels for the image.

If one defines the upper left corner of the digital image (i.e., line one and sample one) to be at scanning coordinates (x_i, z_i), and if one also defines the optical center (origin of the picture coordinate system) to be at scanning coordinates (x_c, z_c), then the x_s and z_s offsets between these two positions in scanning coordinates can be expressed as $\Delta x_s \equiv (x_i - x_c)$ and $\Delta z_s \equiv (z_i - z_c)$. It then follows that the relationship between the digital image and picture coordinate system can be expressed as:

$$x_p = -\Delta x_s + SS(S-1) \tag{3-26}$$

$$z_p = \Delta z_s - SS(L-1) \tag{3-27}$$

or

$$L = 1 - \frac{z_p - \Delta z_s}{SS} \tag{3-28}$$

$$S = 1 + \frac{x_p + \Delta x_s}{SS} . \tag{3-29}$$

4.   THREE DIMENSIONAL STICK FIGURE MODEL.

          A photograph is a two-dimensional representation of the
three-dimensional world. A human is able to understand and interpret
a photograph because he is aware of the 3-D model behind the 2-D
projection. Just as a human can imagine what a photograph would look
like from another camera position, computerized image processing
techniques can synthesize new perspective views if the underlying
3-D model is known. In this project we have used a stick figure
model comprising the edges of planar surfaces in 3-D coordinates as
the representation of the spatial relationships within the scene.

          This section describes the operation of the software that
extracts and manipulates the stick figure model (Section 4.1) and the
mathematical details of the program (Section 4.2.).

4.1   Operation.

          The 3-D stick figure model algorithm is an interactive
program written for the Interactive Digital Image Manipulation
System (IDIMS), which allows an operator to identify planar surfaces
for transformation from one perspective view to another. The algorithm
is composed of two basic modules: the build module and the project
module.

          The build module allows the operator to construct a three-
dimensional stick figure representation of two different types of
features. One feature (type 1) is a simple convex planar surface.
The other feature (type 2) is a rectangular parallelepiped composed
of 6 planar surfaces.

          For type 1 features, the basic scheme is as follows. The
operator places the cursor on any boundary point of the planar surface
and the line and sample coordinates of the point are automatically

4.1   -- Continued.

extracted. He then inputs any one of the (x, y, or z) components of
the spatial coordinate of the point. The inverse perspective equations
are then used to compute the two remaining components of the spatial
coordinate. Spatial coordinates are likewise calculated for at least
two other boundary points (specified consistently in either a clock-
wise or counterclockwise order). As each point is identified it is
marked on the input image and a line is drawn from one point to the
next completely outlining the surface.

After a sufficient number of points have been identified,
the spatial coordinates are used to compute the equation of the best
fit plane (in three dimensions) representing these points. A least
square orthogonal fit method is employed. This method identifies the
coefficients of the planar equation to within a multiplicative factor
of ± 1; that is, it does not identify the direction of the normal to
the plane.

After this step the operator specifies a visibility code
parameter (+1, 0 or -1), which serves a two-fold purpose. It identi-
fies the direction of the normal to the surface, and identifies
whether gray level information is available in the input image (for
transformation of the surface to an output perspective). By defini-
tion the surface normal always points outward from a feature and
generally towards some observer in the spatial coordinate system who
can see the gray level information on the face of the plane. In
terms of the visibility code parameter, a negative value indicates
that the surface points away from the camera and no gray level
information is available, (e.g., the back face of a building).
Another input view is therefore required for this surface. On the
other hand, a non-negative value indicates that the surface normal
generally points towards the camera and gray level information is
visible. A positive parameter value represents an unobstructed view

4.1   -- Continued.

of the surface so that correct gray level information is available.
A zero parameter value indicates that there is gray level information
available, but the surface is partially obstructed by another feature.
(This will lead to an incorrect representation of the surface in the
output perspective and requires another input view or modification of
the gray level information for a correct representation.) Once the
visibility code parameter is specified, the sign of each coefficient
of the equation of the plane is corrected to be consistent with the
direction of the normal in the spatial coordinate system.

The final step is to push the spatial coordinates of the
boundary points back into the plane. This is done by dropping a
perpendicular from the point to the plane and finding the inter-
section point.

For type 2 features, the assumption of a rectangular
parallelepiped allows for the efficient computation of the six sur-
faces by identifying any pair of diagonally opposite corners of the
feature. Figure 4-1 shows a representation of a type 2 feature. The
corners are always identified in the fashion shown. Thus, for
example, the dimensions of the feature are completely identified if
one knows the spatial coordinates for corners 1 and 4.

The basic scheme for this type of feature is as follows.
The operator selects any one of the corners 1, 2, 3, 4, 5 or 7 as a
starting point. In the same manner as for type 1 features, he
positions the cursor at the appropriate line and sample in the image
and specifies any one of the (x, y, or z) components of the coordinates
for that point. The other two components are then calculated. Once
the first corner is specified, he is prompted with a specific sequence
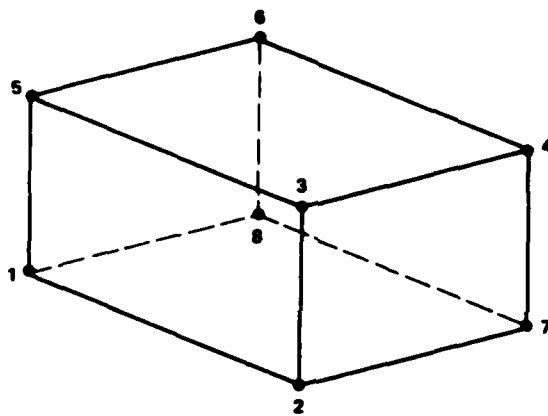of 3 other corners, which will end at the diagonally opposite corner.

Figure 4-1.   Type 2 Feature

4.1   -- Continued.

For example, if corner 7 was initially specified, then the next
corner prompted would be corner 2. The operator then positions the
cursor at the appropriate image coordinate. The algorithm automa-
tically assigns the Z component of corner 2 equal to the Z component
of corner 1 and computes the corresponding X and Y components. The
next prompt will be for corner 3. Once again the operator positions
the cursor and the algorithm assigns independently the X and Y com-
ponents of corner 3 equal to the corresponding components of corner 2.
For each of these components the remaining two spatial components are
calculated and the two resulting sets are then averaged to obtain the
spatial coordinates of corner 3. The last corner prompt is for
corner 5. The operator again places the cursor on the corresponding
point and the algorithm assigns the Z component of corner 5 equal to
the Z component of corner 3. Next the coordinates of the remaining
corners are computed assuming a rectangular parallelepiped configura-
tion. Finally the equations for five surfaces (the base is excluded,
since it can never be observed) are calculated using the least squares
orthogonal fit method, the normals are calculated from assigned
visibility code values for each surface (the two back surfaces are
automatically assigned a visibility code value of -1), and the out-
line of each surface is drawn on the input image.

        The project module allows the stick figure to be projected
from its spatial representation to any two-dimensional output perspec-
tive. Projection can be specified for single points, surfaces,
features or the composite stick figure. The spatial coordinates for
a given surface are projected one at a time using the perspective
equations to the output view, and are then connected by straight
lines in order to outline the surface.

4.2   Mathematical Derivation.

The five key components to the stick figure model algorithm can be summarized as follows:

- Extract spatial coordinates for boundary or corner points.

- Calculate the least square fit planar equation.

- Correct the planar equation for the surface normal.

- Push the boundary points into the plane.

- Project the 3-D stick figure model to 2-D output perspective.

The mathematical analysis of each of these five key components is now described below.

4.2.1   Extract Spatial Coordinates.

The extraction of spatial coordinates (x, y, z) corresponding to a given set of digital image coordinates first requires the conversion from a set of digital image coordinates to a set of picture coordinates, followed by an application of the inverse perspective equations. The former can be obtained by a straightforward application of equations (3-26) and (3-27). The latter is more complex, because only two picture coordinates are input, and there are three inverse perspective equations (2-20), (2-21) and (2-22) for each of the spatial coordinates. Thus, in order to obtain the inverse perspective transformation, at least one spatial coordinate must also be specified. With this in mind we shall treat the three cases separately,

4.2.1 -- Continued.

each one pertaining to a different initially specified component of
the spatial coordinate system.

Case A

If z is initially specified, then one can solve for λ from
equation (2-22). Thus

$$\lambda = \frac{z - z_o}{M_{31}x_p + M_{32}z_p + M_{33}f} \quad .$$

Substituting this equation back into equations 20 and 21 to eliminate
the free parameter leads to

$$x = x_o + (z-z_o) \left[\frac{M_{11}x_p + M_{12}z_p + M_{13}f}{M_{31}x_p + M_{32}z_p + M_{33}f} \right] \qquad (4-1)$$

$$y = y_o + (z-z_o) \left[\frac{M_{21}x_p + M_{22}z_p + M_{23}f}{M_{31}x_p + M_{32}z_p + M_{33}f} \right] \qquad (4-2)$$

Case B

In a similar fashion, if y is specified initially, then
one obtains

$$x = x_o + (y-y_o) \left[\frac{M_{11}x_p + M_{12}z_p + M_{13}f}{M_{21}x_p + M_{22}z_p + M_{23}f} \right] \qquad (4-3)$$

$$z = z_o + (y-y_o) \left[\frac{M_{31}x_p + M_{32}z_p + M_{33}f}{M_{21}x_p + M_{22}z_p + M_{23}f} \right] \qquad (4-4)$$

4.2.1   -- Continued.

Case C

In the same manner, if x is specified initially, then one obtains

$$y = y_o + (x-x_o) \left[\frac{M_{21}x_p + M_{22}z_p + M_{23}f}{M_{11}x_p + M_{12}z_p + M_{13}f} \right] \qquad (4-5)$$

$$z = z_o + (x-x_o) \left[\frac{M_{31}x_p + M_{32}z_p + M_{33}f}{M_{11}x_p + M_{12}z_p + M_{13}f} \right] \qquad (4-6)$$

These three pairs of equations completely describe the inverse perspective equations as applied to this problem.

4.2.2   Least Square Orthogonal Planar Fit.

While the method of fitting curves to data in the least squares sense is a very familiar one, most published analyses and applications define only one component of the distance of each point from the curve as the error whose sum of squares is to be minimized. In effect, one component of the distance is defined as the dependent one, while the others are defined as independent. The choice of dependent variable affects the solution obtained. For example fitting a straight line (Ax + By +C = 0) to a set of points produces one solution if one minimizes the sum of the squares of the errors Δy and another solution by minimizing the Δx sum of squares. The following analysis presents a method for obtaining the best planar fit to a set of points, where best is defined as minimizing the sum of the squares of the orthogonal errors.

4.2.2   -- Continued.

          Let the general equation for the plane that is to be fit to
the points be:

$$Ax + By + Cz + D = 0 \qquad (4-7)$$

Then, the orthogonal distance of a point (x_i, y_i, z_i) from this plane
is given by:

$$R_i = \frac{(Ax_i + By_i + Cz_i + D)}{(A^2 + B^2 + C^2)^{1/2}} \qquad (4-8)$$

          Thus, the function that is to be minimized is:

$$E = \sum_{i=1}^{N} R_i^2 = \sum_{i=1}^{N} \frac{(Ax_i + By_i + Cz_i + D)^2}{A^2 + B^2 + C^2} \qquad (4-9)$$

Using the standard technique of taking first partial derivatives of
(4-9) with respect to the coefficients (A,B,C,D) we have (after
dividing through by $2(A^2+B^2+C^2)$):

$$\sum_{i=1}^{N} \left[(Ax_i+By_i+Cz_i+D)x_i - \frac{A(Ax_i+By_i+Cz_i+D)^2}{A^2+B^2+C^2} \right] = 0 \qquad (4-10)$$

$$\sum_{i=1}^{N} \left[(Ax_i+By_i+Cz_i+D)y_i - \frac{B(Ax_i+By_i+Cz_i+D)^2}{A^2+B^2+C^2} \right] = 0 \qquad (4-11)$$

$$\sum_{i=1}^{N} \left[(Ax_i+By_i+Cz_i+D)z_i - \frac{C(Ax_i+By_i+Cz_i+D)^2}{A^2+B^2+C^2} \right] = 0 \qquad (4-12)$$

4.2.2 — Continued.

$$\sum_{i=1}^{N} (Ax_i + By_i + Cz_i + D) = 0 \qquad (4-13)$$

Then, define a new parameter, γ, as:

$$\gamma = \sum_{i=1}^{N} \left[\frac{(Ax_i + By_i + Cz_i + D)^2}{A^2 + B^2 + C^2} \right].\qquad (4-14)$$

Equations (4-10) through (4-12) now become:

$$A\left(\sum_{i=1}^{N} x_i^2 - \gamma\right) + B\sum_{i=1}^{N} x_i y_i + C\sum_{i=1}^{N} x_i z_i + D\sum_{i=1}^{N} x_i = 0 \qquad (4-15)$$

$$A\sum_{i=1}^{N} x_i y_i + B\left(\sum_{i=1}^{N} y_i^2 - \gamma\right) + C\sum_{i=1}^{N} y_i z_i + D\sum_{i=1}^{N} y_i = 0 \qquad (4-16)$$

$$A\sum_{i=1}^{N} x_i z_i + B\sum_{i=1}^{N} y_i z_i + C\left(\sum_{i=1}^{N} z_i^2 - \gamma\right) + D\sum_{i=1}^{N} z_i = 0 \quad, \qquad (4-17)$$

and equation (4-13) can be solved for D as:

$$D = -A\bar{x} - B\bar{y} - C\bar{z} \quad, \qquad (4-18)$$

where the bar over the variable implies the mean value (eg., $\bar{x} \equiv \frac{1}{N}\sum_{i=1}^{N} x_i$).

Substituting (4-18) into the previous three equations yields (in matrix form):

4.2.2 ___ -- <u>Continued.</u>

$$
\begin{bmatrix}
(\sum_{i=1}^{N} x_i^2 - N\bar{x}^2 - \gamma) & (\sum_{i=1}^{N} \bar{x}_i \bar{y}_i - N\bar{x}\bar{y}) & (\sum_{i=1}^{N} \bar{x}_i \bar{z}_i - N\bar{x}\bar{z}) \\
(\sum_{i=1}^{N} x_i y_i - N\bar{x}\bar{y}) & (\sum_{i=1}^{N} y_i^2 - N\bar{y}^2 - \gamma) & (\sum_{i=1}^{N} y_i z_i - N\bar{y}\bar{z}) \\
(\sum_{i=1}^{N} x_i z_i - N\bar{x}\bar{z}) & (\sum_{i=1}^{N} \bar{y}_i \bar{z}_i - N\bar{y}\bar{z}) & (\sum_{i=1}^{N} \bar{z}_i^2 - N\bar{z}^2 - \gamma)
\end{bmatrix}
\begin{bmatrix} A \\ B \\ C \end{bmatrix}
=
\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}
\quad (4\text{-}19)
$$

But, note that $\sum_{i=1}^{N} \bar{x}_i^2 - N\bar{x}^2 = \sum_{i=1}^{N} (x_i - \bar{x})^2 \equiv \sum_{i=1}^{N} \tilde{x}_i^2$

and $\sum_{i=1}^{N} x_i y_i - N\bar{x}\bar{y} = \sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y}) \equiv \sum_{i=1}^{N} \tilde{x}_i \tilde{y}_i$.

Thus, (4-19) may be rewritten as:

$$
\begin{bmatrix}
(\sum_{i=1}^{N} \tilde{x}_i^2 - \gamma) & (\sum_{i=1}^{N} \tilde{x}_i \tilde{y}_i) & (\sum_{i=1}^{N} \tilde{x}_i \tilde{z}_i) \\
(\sum_{i=1}^{N} \tilde{x}_i \tilde{y}_i) & (\sum_{i=1}^{N} \tilde{y}_i^2 - \gamma) & (\sum_{i=1}^{N} \tilde{y}_i \tilde{z}_i) \\
(\sum_{i=1}^{N} \tilde{x}_i \tilde{z}_i) & (\sum_{i=1}^{N} \tilde{y}_i \tilde{z}_i) & (\sum_{i=1}^{N} \tilde{z}_i^2 - \gamma)
\end{bmatrix}
\begin{bmatrix} A \\ B \\ C \end{bmatrix}
=
\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}
\quad (4\text{-}20)
$$

or

$$
\begin{bmatrix}
\tilde{\alpha}_{11} & \tilde{\alpha}_{12} & \tilde{\alpha}_{13} \\
\tilde{\alpha}_{21} & \tilde{\alpha}_{22} & \tilde{\alpha}_{23} \\
\tilde{\alpha}_{31} & \tilde{\alpha}_{32} & \tilde{\alpha}_{33}
\end{bmatrix}
\begin{bmatrix} A \\ B \\ C \end{bmatrix}
=
\begin{bmatrix}
(\alpha_{11} - \gamma) & \alpha_{12} & \alpha_{13} \\
\alpha_{21} & (\alpha_{22} - \gamma) & \alpha_{23} \\
\alpha_{31} & \alpha_{32} & (\alpha_{33} - \gamma)
\end{bmatrix}
\begin{bmatrix} A \\ B \\ C \end{bmatrix}
=
\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}
\quad (4\text{-}21)
$$

4.2.2   -- Continued.

Thus, in order for equation (4-19) to have a non-trivial solution
(A≠0, B≠0, C≠0), the matrix must be singular (i.e., the determinant
must = 0). This can be accomplished through a suitable choice of γ,
which is an eigenvalue of the matrix. Recall that γ was defined in
terms of the coefficients A, B, C, and D, and is in fact, the very
function that is to be minimized. The determinant of the matrix
in (4-20) is a third order polynomial in γ:

$$\gamma^3 + p\gamma^2 + q\gamma + r = 0 \tag{4-22}$$

where $p = -(\alpha_{11} + \alpha_{22} + \alpha_{23})$

$$q = \alpha_{11}\alpha_{22} + \alpha_{11}\alpha_{33} + \alpha_{22}\alpha_{33} - \alpha_{12}^2 - \alpha_{13}^2 - \alpha_{23}^2$$

$$r = \alpha_{11}\alpha_{23}^2 + \alpha_{22}\alpha_{13}^2 + \alpha_{33}\alpha_{12}^2 - 2\alpha_{12}\alpha_{13}\alpha_{23} - \alpha_{11}\alpha_{22}\alpha_{33}$$

Equation (4-22) has three roots, all of which must be real, since they
are eigenvalues of the symmetric matrix in (4-20). The parameter γ,
as defined by (4-14) is the function that is to be minimized, so the
only valid (non-extraneous) solution is the smallest non-negative
eigenvalue. Note that if r = 0, then the solution desired is $\gamma = 0$.

If $r \neq 0$, then the three roots of (4-22) may be found from:

$$\gamma_1 = 2\sqrt{\frac{-a}{3}} \cos(\beta/3) - p/3$$

$$\gamma_2 = 2\sqrt{\frac{-a}{3}} \cos(\beta/3 + 120°) - p/3 \tag{4-23}$$

$$\gamma_3 = 2\sqrt{\frac{-a}{3}} \cos(\beta/3 + 240°) - p/3$$

4.2.2   -- Continued.

where:

$$a = q - p^2/3 \qquad\qquad (4\text{-}24)$$

and,

$$p = \text{principal value of } \cos^{-1}\left[\frac{-b}{2\sqrt{-a^3/27}}\right] \qquad (4\text{-}25)$$

$$b = r - pq/3 + (2/27)p^3 \qquad\qquad (4\text{-}26)$$

Once the smallest non-negative eigenvalue is found, it is used to determine the elements $\tilde{\alpha}_{ij}$ of the matrix in (4-21). Then, the coefficients of the desired equation (4-7) may be found from:

$$
\begin{aligned}
A &= k(\tilde{\alpha}_{i2}\tilde{\alpha}_{j3} - \tilde{\alpha}_{j2}\tilde{\alpha}_{i3}) \\
B &= k(\tilde{\alpha}_{j1}\tilde{\alpha}_{i3} - \tilde{\alpha}_{i1}\tilde{\alpha}_{j3}) \\
C &= k(\tilde{\alpha}_{i1}\tilde{\alpha}_{j2} - \tilde{\alpha}_{j1}\tilde{\alpha}_{i2})
\end{aligned}
\qquad (4\text{-}27)
$$

Any two rows, i and j from (4-21) are suitable provided that the choice does not result in all parenthetical factors = 0. In the above expressions k is a proportionality constant that is arbitrarily chosen to be $k = \pm(A^2+B^2+C^2)^{1/2}$ in order to yield conveniently sized coefficients for equation (4-7). Finally C can then be uniquely determined from equation (4-18):

$$D = -A\bar{x} - B\bar{y} - C\bar{z} \qquad\qquad (4\text{-}28)$$

Thus, the least squares solution for orthogonal errors has been determined for a linear function of the variables x_i, y_i, and z_i.

4.2.3   Direction of the Surface Normal.

In the previous section, the coefficients describing the least squares method for fitting plane were described. The first three components (A,B,C) also comprise the vector (in spatial coordinates), which is normal to the plane. The inclusion of a dual-valued scaling constant k allows the freedom to pick one of two opposing directions for this normal, so that it may extend from the side of the plane that contains gray level information. The selection of the appropriate sign to use is determined from the visibility code parameter in the following manner.

Recall that a non-negative visibility code indicates that the important gray level detail of a surface should be visible to the camera and that a negative visibility code indicates that the gray level detail is not visible. Now, since the feature is defined in the input scene (it is not behind the camera), a sufficient test to use to identify the sign for the surface coefficients can be derived using the visibility code. Start with $k > 0$. If the visibility code is non-negative, then the camera must be on the same side of the surface as the gray level detail. If the position of the camera (x_o, y_o, z_o) is inserted into the expression for the plane, $Ax + By + Cz + D$, then a positive number should result. If it does not, then choose $k < 0$. Similarly, if the visibility code is negative, then the camera is on the back side of the plane, i.e., the opposite side from that with the gray level information. In this case, if the camera position is inserted into the expression for the plane, a negative number should result. If it does not, then choose $k < 0$.

4.2.4   Push Boundary Points into the Plane.

The spatial boundary points used to extract the least square orthogonal fit plane do not necessarily lie in the plane. In order to ensure a consistent outline of the surface in the output perspective,

4.2.4   -- Continued.

it is necessary to push these points back into the plane. This is
done by dropping a perpendicular line from each point to the plane
and finding the intersection of this perpendicular line with the plane.

Consider a point $p_1 = (x_1, y_1, z_1)$ and a plane $Ax + By + Cz + D = 0$. A line that passes through this point that is perpendicular to the plane (parallel to the normal vector $\vec{n} = A\hat{\imath} + B\hat{\jmath} + C\hat{k}$) can be expressed as

$$\frac{x-x_1}{A} = \frac{y-y_1}{B} = \frac{z-z_1}{C} \qquad (4-29)$$

If the point of intersection of this line with the plane is called $p_o = (x_o, y_o, z_o)$ then, point p_o must satisfy equation (4-29), so that

$$\frac{x_o-x_1}{A} = \frac{y_o-y_1}{B} = \frac{z_o-z_1}{C} \qquad (4-30)$$

Point p_o must also satisfy the planar equation, so that

$$Ax_o + By_o + Cz_o + D = 0 \qquad (4-31)$$

One may now solve for (By_o) and for (Cz_o) in terms of x_o using equation (4-30).

$$By_o = \frac{B^2(x_o-x_1)}{A} + By_1 \qquad (4-32)$$

and

$$Cz_o = \frac{C^2(x_o-x_1)}{A} + Cz_1 \qquad (4-33)$$

4-15

4.2.4   -- Continued.

Substitution of equations (4-32) and (4-33) into equation (4-31) leads to

$$x_O = \frac{-AD + (B^2+C^2)x_1 - ABy_1 - ACz_1}{A^2+B^2+C^2} \quad . \qquad (4-34)$$

In a similar manner one also obtains

$$y_O = \frac{-BD + (A^2+C^2)y_1 - BAx_1 - BCz_1}{A^2+B^2+C^2} \qquad (4-35)$$

and

$$z_O = \frac{-CD + (A^2+B^2)z_1 - CAx_1 - CBy_1}{A^2+B^2+C^2} \quad . \qquad (4-36)$$

4.2.5   3-D to 2-D Projection.

The projection of a set of spatial points to a given output perspective view is described by the perspective equations (2-9) and (2-10).

5. GRAY LEVEL TRANSFORMATION.

Gray level transformation is the process whereby the pixel by pixel gray level or intensity information is transferred from the original digital imagery to the output perspective view. A few papers have been published wherein gray level texture prototypes are transferred via perspective transforms to generate synthetic images[6,7], but the specific problem of transferring actual gray level information for real scenes has not been previously addressed.

For each planar surface imaged, the input and output image models and the planar surface equation in the stick figure model are used to define the mathematical transformation (Section 5.1). All of the transformed surfaces making up the output image are combined in the proper relationship to one another (Section 5.2).

5.1 Generation and Application of the Gray Level Transformation.

Section 2.2 explained that the two camera perspective equations depend on information about the scene being imaged as well as the image models for the two cameras. We have made the assumption that the urban scenes we work with can be modeled as a collection of planar surfaces. The stick figure model discussed in Section 4 contains the information that defines each planar surface: the equation of the plane and the vertices that define the plane's polygonal boundary.

The stick figure model information for a given plane is used to supply the parameters relating the input and output image coordinate systems in equations (2-32) and (2-33). The auxiliary relationships (equations (3-36) and (3-27)) between image coordinates and pixels are used so that the transformation between the input and output coordinates can be computed:

5.1   -- Continued.

$$S = \frac{P_1 S' + P_2 L' + P_3}{Q_1 S' + Q_2 L' + Q_3} \qquad (5\text{-}1)$$

$$L = \frac{R_1 S' + R_2 L' + R_3}{Q_1 S' + Q_2 L' + Q_3} \qquad (5\text{-}2)$$

where

S = sample number in input image

L = line number in input image

S´ = sample number in output image

L´ = line number in output image

P_i, Q_i, R_i = parameters that depend upon the image models and plane equation but not upon S, L, S´ and L´.

The fact that the parameters P_i, Q_i, and R_i do not depend upon the pixel locations means that equations (5-1) and (5-2), which must be evaluated at each pixel, are quick computationally.

The spatial coordinates of the planar surface vertices are projected to the output image coordinate system using equations (2-9) and (2-10), and converted to line and sample by equations (3-28) and (3-29). Equations (5-1) and (5-2) are evaluated at each output image pixel within the polygon defined by the projected vertices and represent the input image pixel location from which the gray level is to be taken. In general, L and S fall between pixels, so bilinear interpolation is done to get the gray level:

5.1   -- Continued.

Gray level = $(1-S_F)[(1-L_F)I(L_I,S_I) + L_FI(L_I+1,S_I)]$

$+S_F[(1-L_F)I(L_I,S_I+1) + L_FI(L_I+1,S_I+1)]$

where

S_I = the integer part of S

S_F = the fractional part of S

L_I = the integer part of L

L_F = the fractional part of L

I = the gray levels of the input image.

Evaluating equations (5-1), (5-2), and (5-3) for each pixel of the output image that falls within the polygon defined by the projected vertices completes the gray level transformation for the planar surface. This process is repeated for every plane that is contained in the output image.

5.2   Combining the Transformed Planes.

The planes that make up the input image, after gray level transformation, make up the output image. Two significant processes must be performed: missing surface addition and hidden surface removal. The former problem, filling in the areas that can be seen from the output perspective but not from the input perspective, has not been addressed yet in this study.

The latter problem consists of removing the surfaces or parts of surfaces that fall behind other surfaces as seen from the output perspective. A related problem, removal of the hidden surfaces in graphics manipulations, has received substantial attention in recent years[8].

5.2   -- Continued.

      These techniques are not completely general in their ability to solve the hidden surface problem and can involve very intricate logical tests. Up to this time, an essentially manual approach has been used in this project. The analyst inspects the input image and the stick figure model to determine which surfaces should have priority over the other surfaces on the basis of the proximity to the output camera. The analyst orders the surfaces so that the later ones are closer to the camera. The surfaces are transformed in this order and written to a single output image file. Each surface is written over any surfaces that may have been written previously at the same location on a pixel-by-pixel basis. Thus, if the analyst orders the surfaces properly, the hidden surface problem is handled by an overlay process.

      Work is currently being done to make the hidden surface removal process more automatic. The concept being implemented includes the following steps to eliminate entire surfaces from transformation:

- A surface will be eliminated if its boundary falls outside the output image (past one of the edges).

- A surface will be eliminated if its boundary points are behind the output camera.

- A surface will be eliminated if the output camera is on the wrong side of the surface to see the gray level information.

In addition, a pixel-by-pixel test will be done to determine which of several competing surfaces should provide the gray level information at each point of the output image. The distance from the camera to

5.2   -- Continued.

the surfaces (along the ray through the pixel) will be computed. The
closest surface to the camera at each pixel will provide the gray
level. These automatic tests should be sufficient to supplant the
manual techniques, thereby reducing the time spent on the process
and eliminating errors.

6._____EXAMPLES.

6.1_____Original Photography.

The aerial photography was taken with an RC-10 camera with a 12-inch focal length at approximately 3000 foot altitude. Nine-inch format Plus-X aerographic film was used.

Figures 6-1 and 6-2 are examples of the original imagery. The oblique photographs, such as Figure 6-1, were used to produce the output imagery. The vertical photographs, such as Figure 6-2, provided some location information, although the U.S.G.S. 7 1/2-minute quadrangle maps supplied most of this data.

6.2_____Stick Figure Models.

The stick figure models were generated using the computer program described in Section 4, which allows a user to interactively select image points to define a stick figure. Figure 6-3 shows a digitized image and the stick figure built for it. Note that the automobiles have been edited out of the scene in Figure 6-3(b). There are five surfaces in the stick figure model:

        a.   The sunny side of the building
        b.   The shady side
        c.   The roof
        d.   The east-west street
        e.   The north-south street.

Table 6-1 shows the equations and boundaries of the five planes.
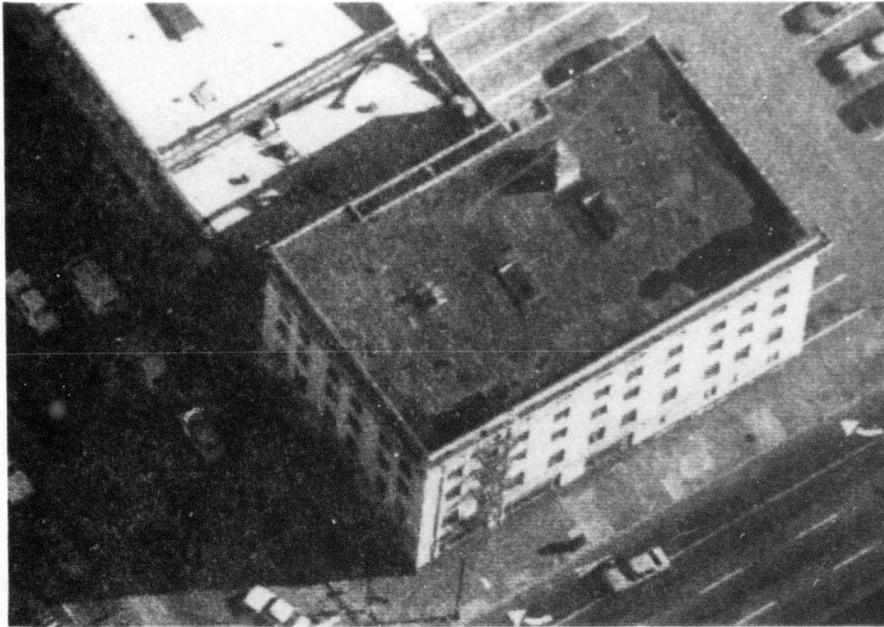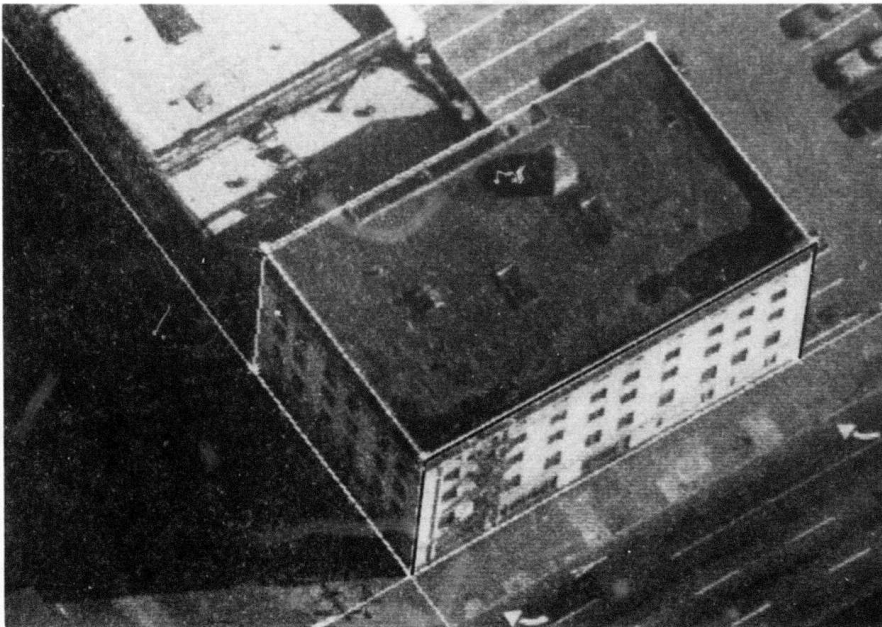
Figure 6-1. Oblique Aerial Photograph
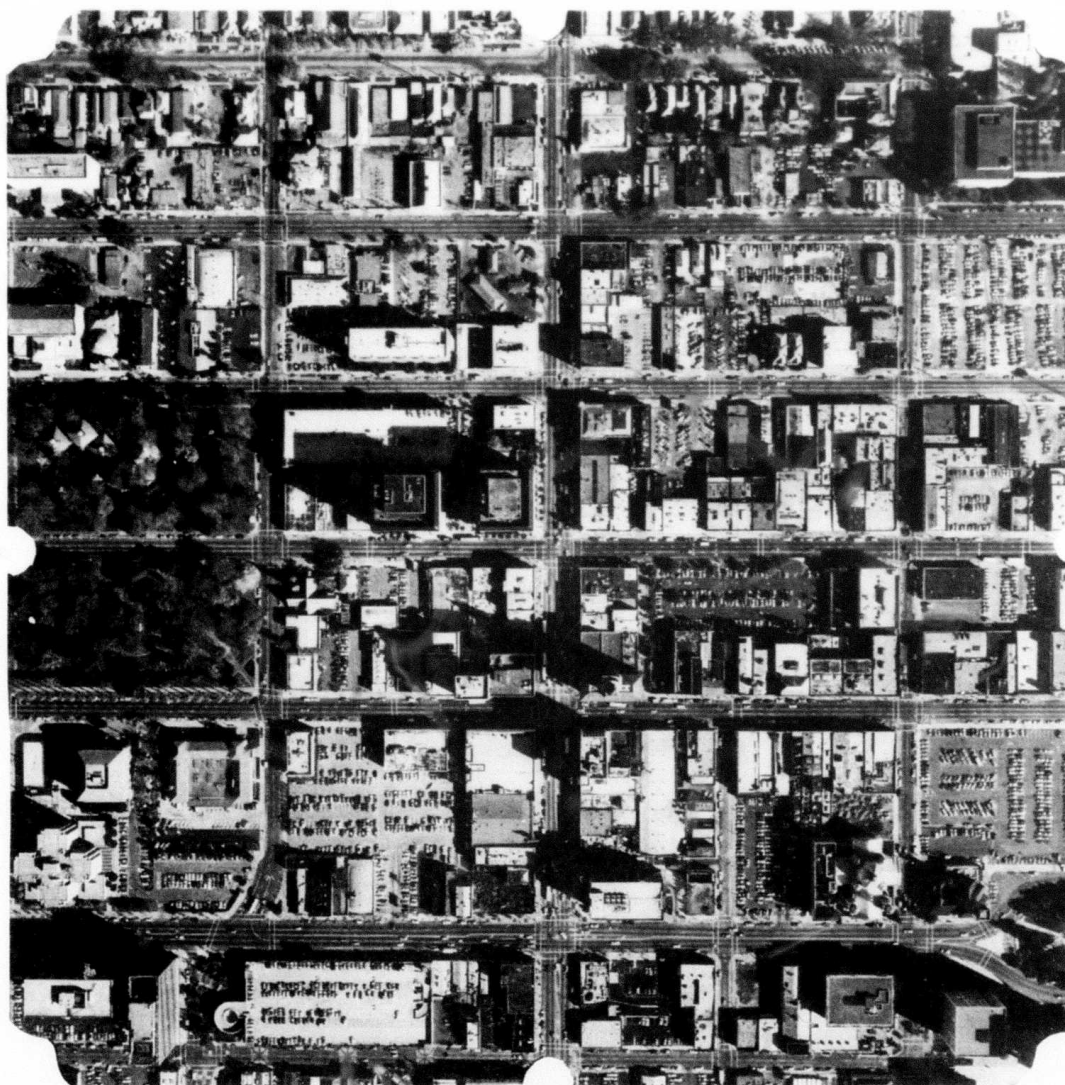
Figure 6-2. Vertical Overhead Photograph

a. Digitized scene

b. Stick figure model

Figure 6-3. Digitized Scene and Stick Figure Model

Table 6-1. Stick Figure Parameters

Surface	Equation*	Boundary points*		
		x	y	z
Sunny side	$-.05x-1.00y-.02z = 410.00$	a. -807.7	-404.6	-77.1
		b. -807.1	-405.8	-12.3
		c. -713.2	-406.2	-12.3
		d. -713.1	-405.0	-77.1
Shady side	$-1.00x + .01z = 807.28$	a. -807.8	-404.6	-77.1
		b. -807.1	-405.8	-12.3
		c. -807.1	-346.4	-12.3
		d. -807.8	-345.7	-77.1
Roof	$z = -12.3$	a. -807.0	-346.2	-12.3
		b. -807.2	-405.3	-12.3
		c. -713.2	-406.4	-12.3
		d. -714.0	-347.4	-12.3
East-west street	$z = -77.10$	a. -808.3	-243.0	-77.1
		b. -807.9	-431.6	-77.1
		c. -891.7	-375.7	-77.1
North-south street	$z = -77.10$	a. -848.8	-404.4	-77.1
		b. -690.8	-405.1	-77.1
		c. -739.2	-477.4	-77.1

*Units are feet. The x-axis is parallel to the sunny side, the y-axis is parallel to the shady side, and the z-axis is vertical.

6.3 Synthesized Perspective Views.

The stick figure model and the digitized image were used to generate a sequence of perspective views with the computer program described in Section 5. The outputs were sharpened using digital image processing techniques.

Figures 6-4 thru 6-7 simulate a camera 100 feet across the street from the sunny face of the building, with the optical axis aimed at a point on the building 10 feet above street level. In Figure 6-4, the camera is 10 feet above street level and has no tilt. In Figure 6-5, the camera is 70 feet above the street, nearly at the height of the roof and has moderate tilt. In Figure 6-6, the camera is 170 feet above the street, tilted significantly downwards. Figure 6-7 simulates what a camera would see if it could be placed 70 feet below street level.

Figures 6-8 thru 6-11 simulate a camera 70 feet diagonally opposite the corner of the building, going through the same sequence of heights and tilts as the previous set.

Even though a human may tilt his head in viewing a tall building, the untilted views in these figures are more comfortable to view than the tilted ones. Figures 6-12 thru 6-14 illustrate three views that an observer might see as he walks down the sidewalk opposite the shady side of the street. A 45° pan angle was chosen as a reasonable approximation to the forward viewing of the observer.
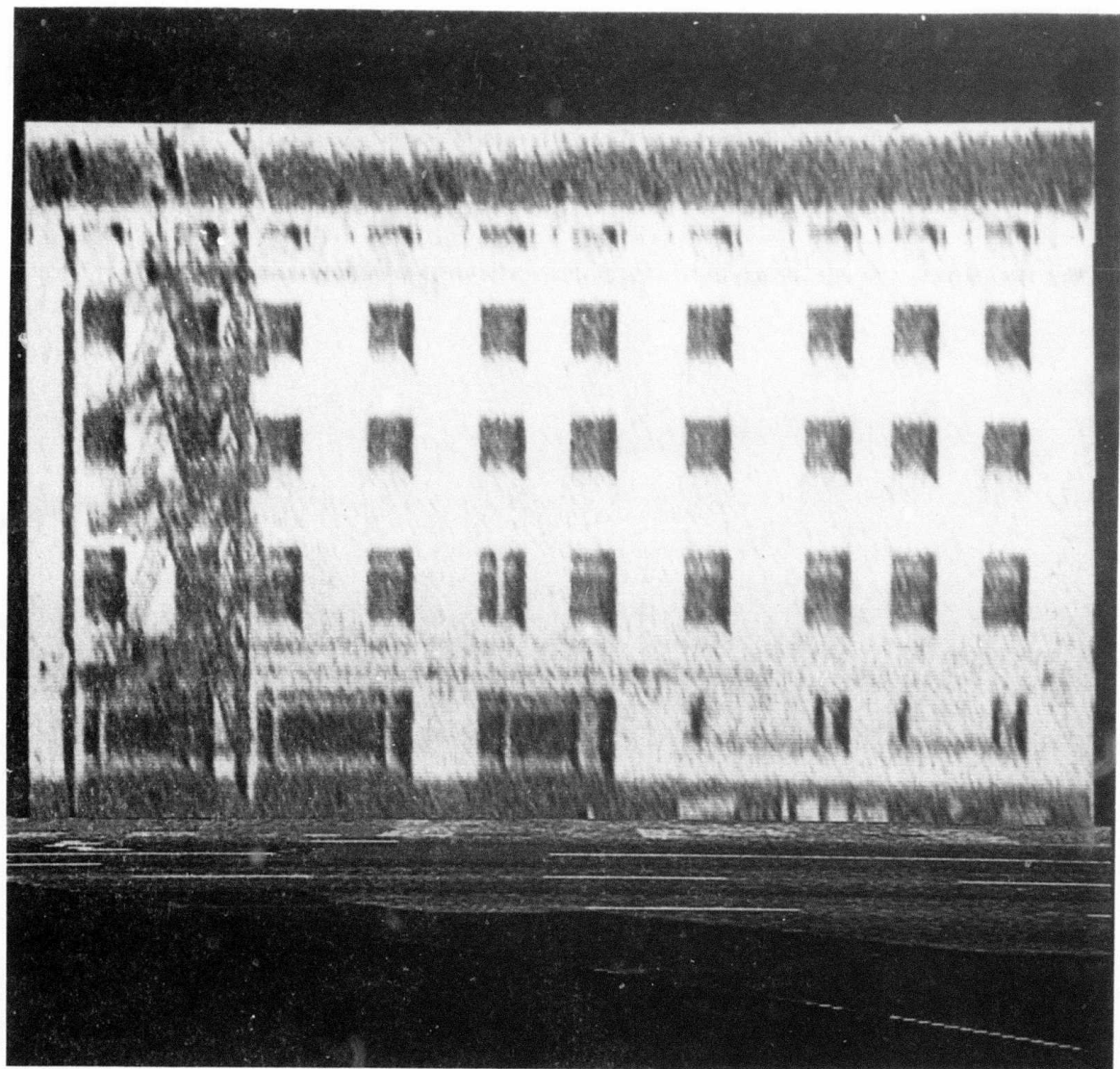
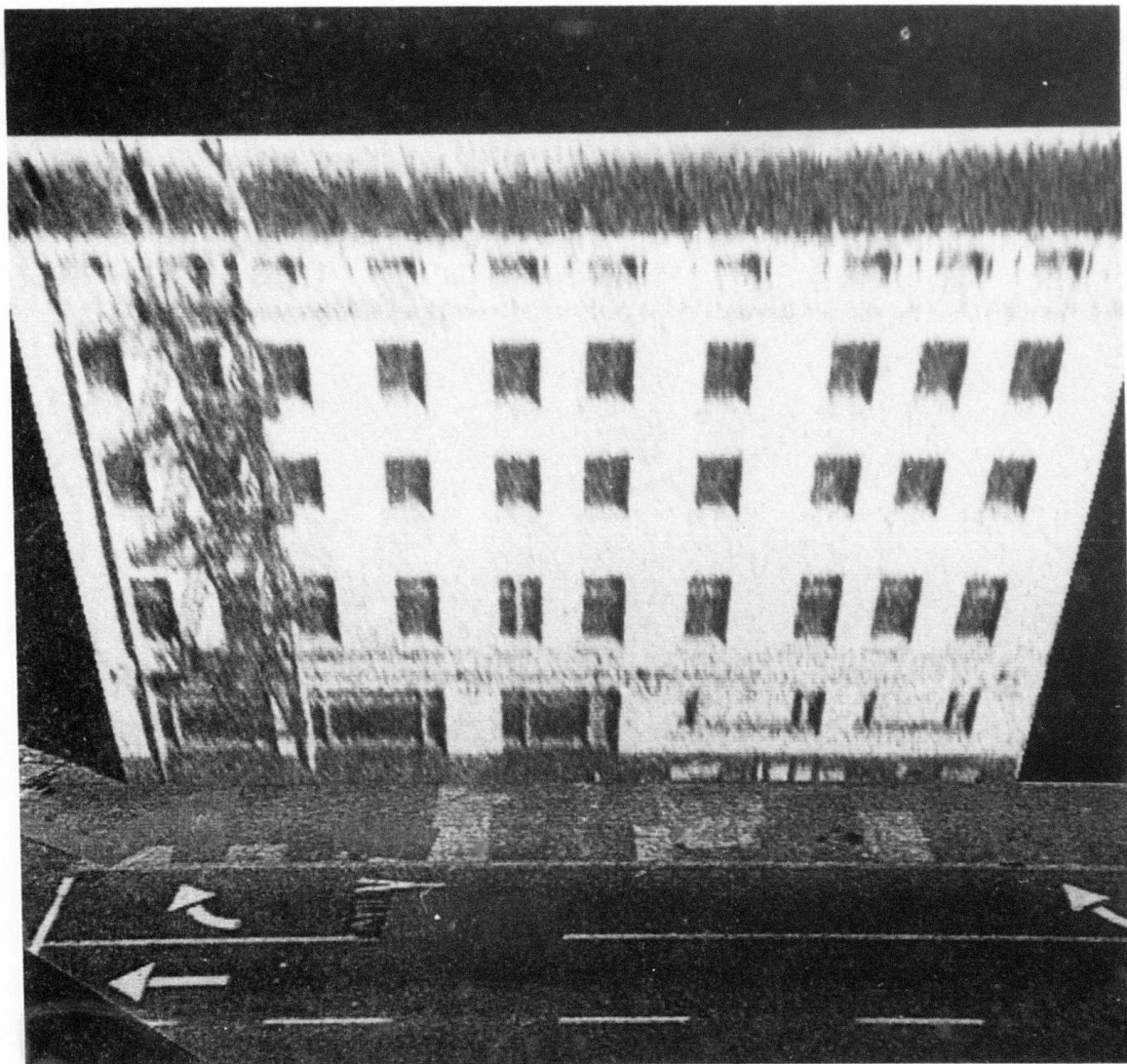Figure 6-4.   Face-on View, 10 Foot Altitude, No Tilt

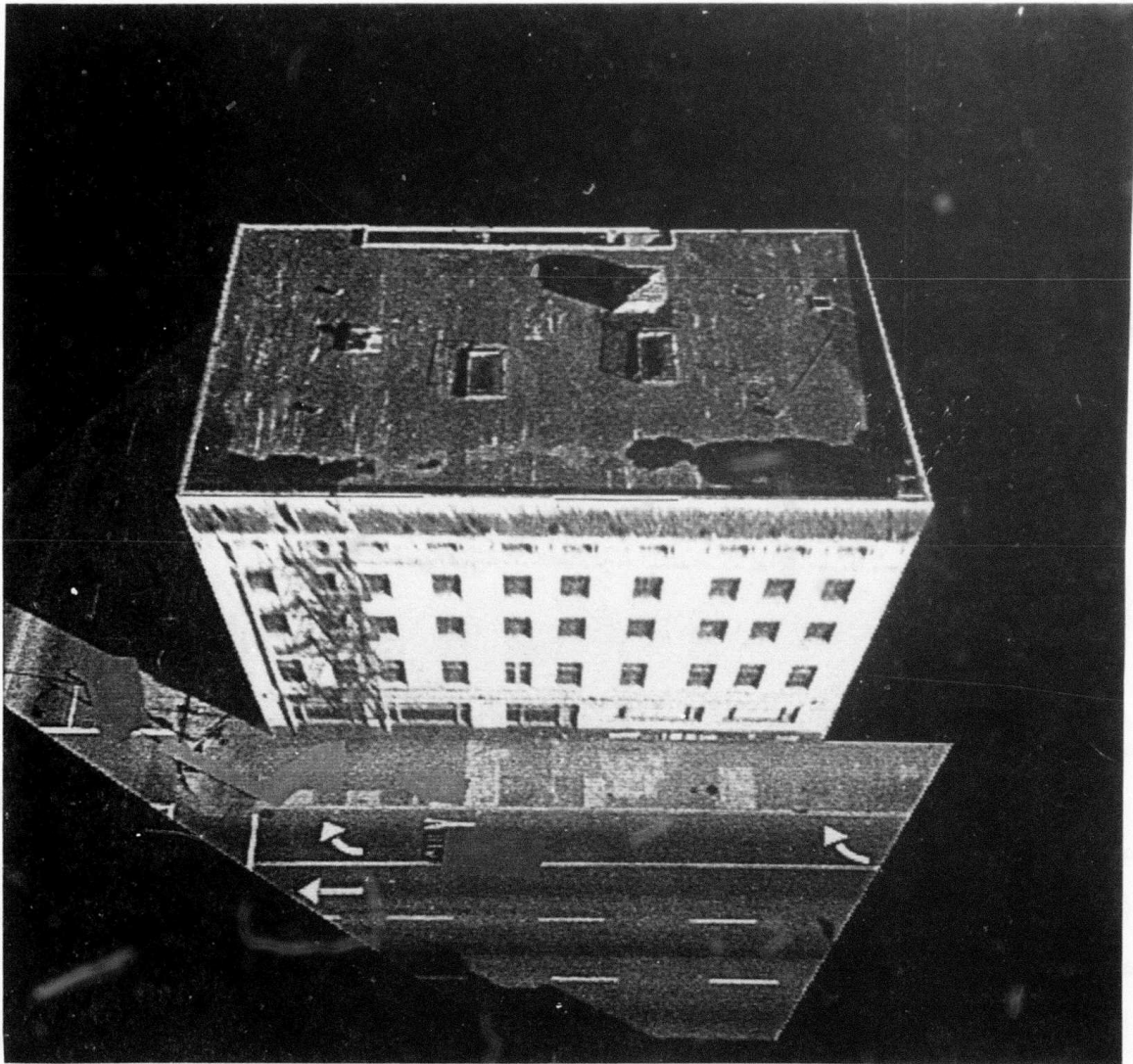Figure 6-5.   Face-on View, 70 Foot Altitude, Moderate Tilt

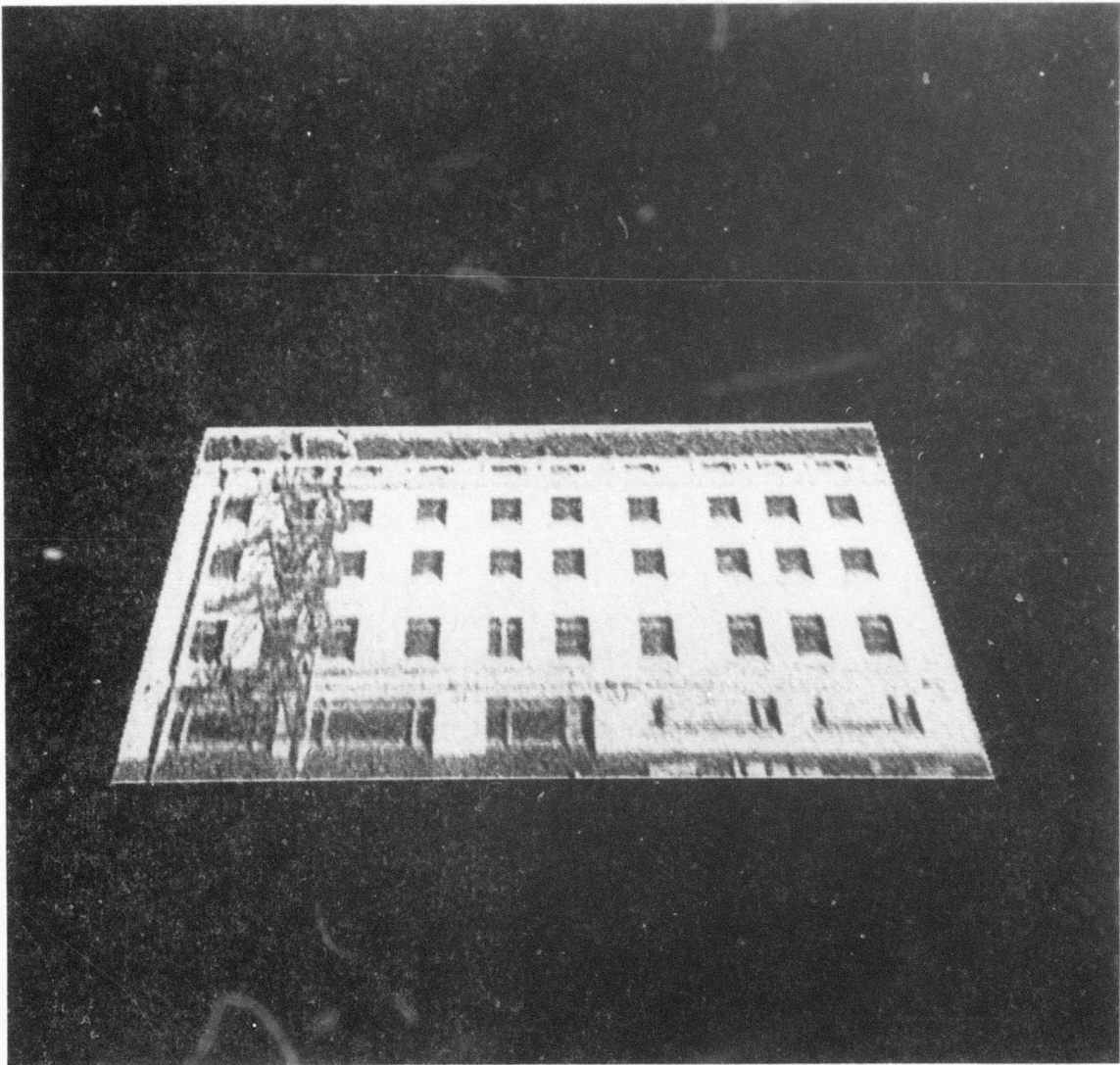Figure 6-6.   Face-on View, 170 Foot Altitude, Significant Tilt

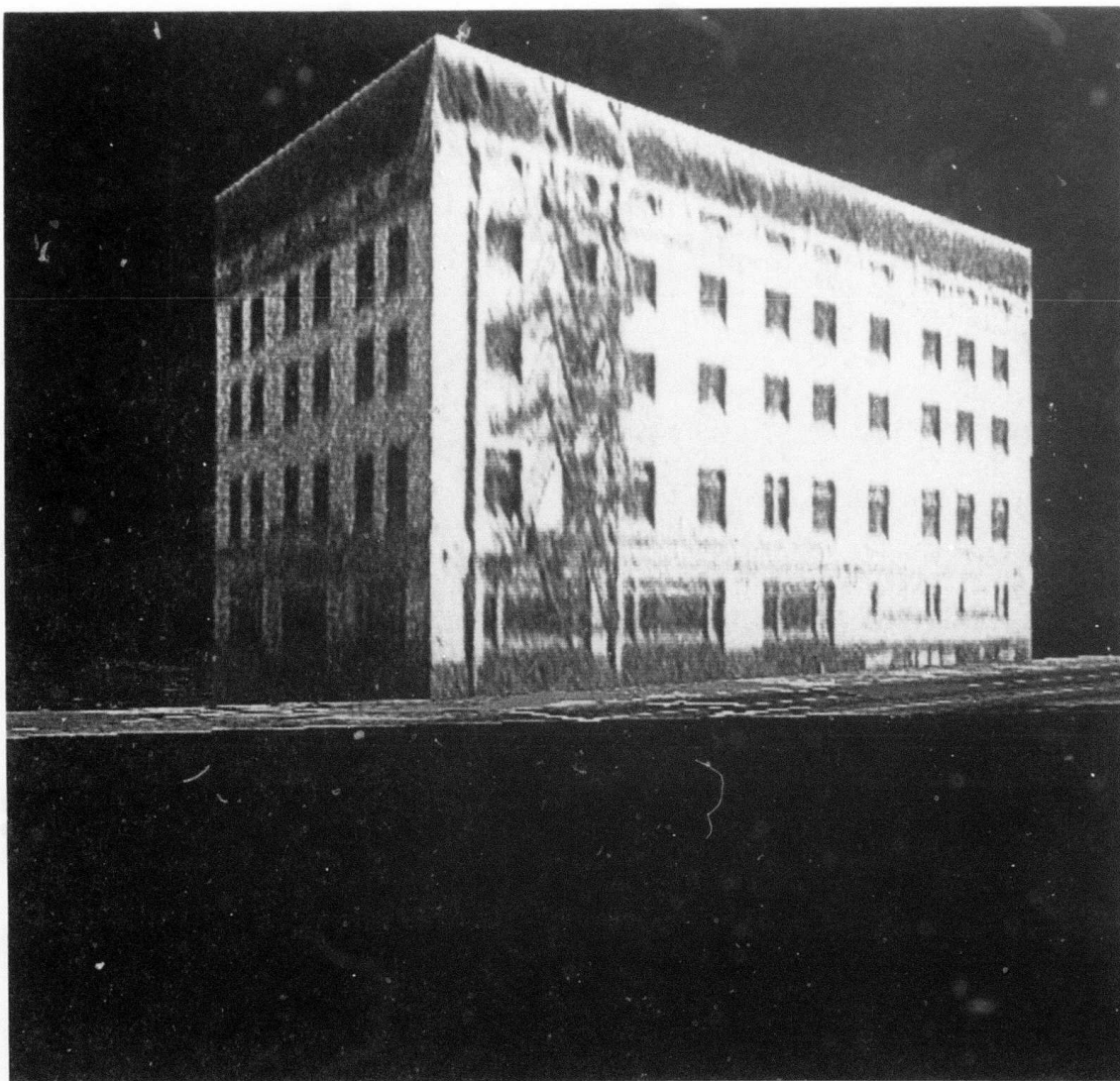Figure 6-7.   Face-on View, Below Ground Level, Upward Tilt

Figure 6-8. Diagonal View, 10 Foot Altitude, No Tilt

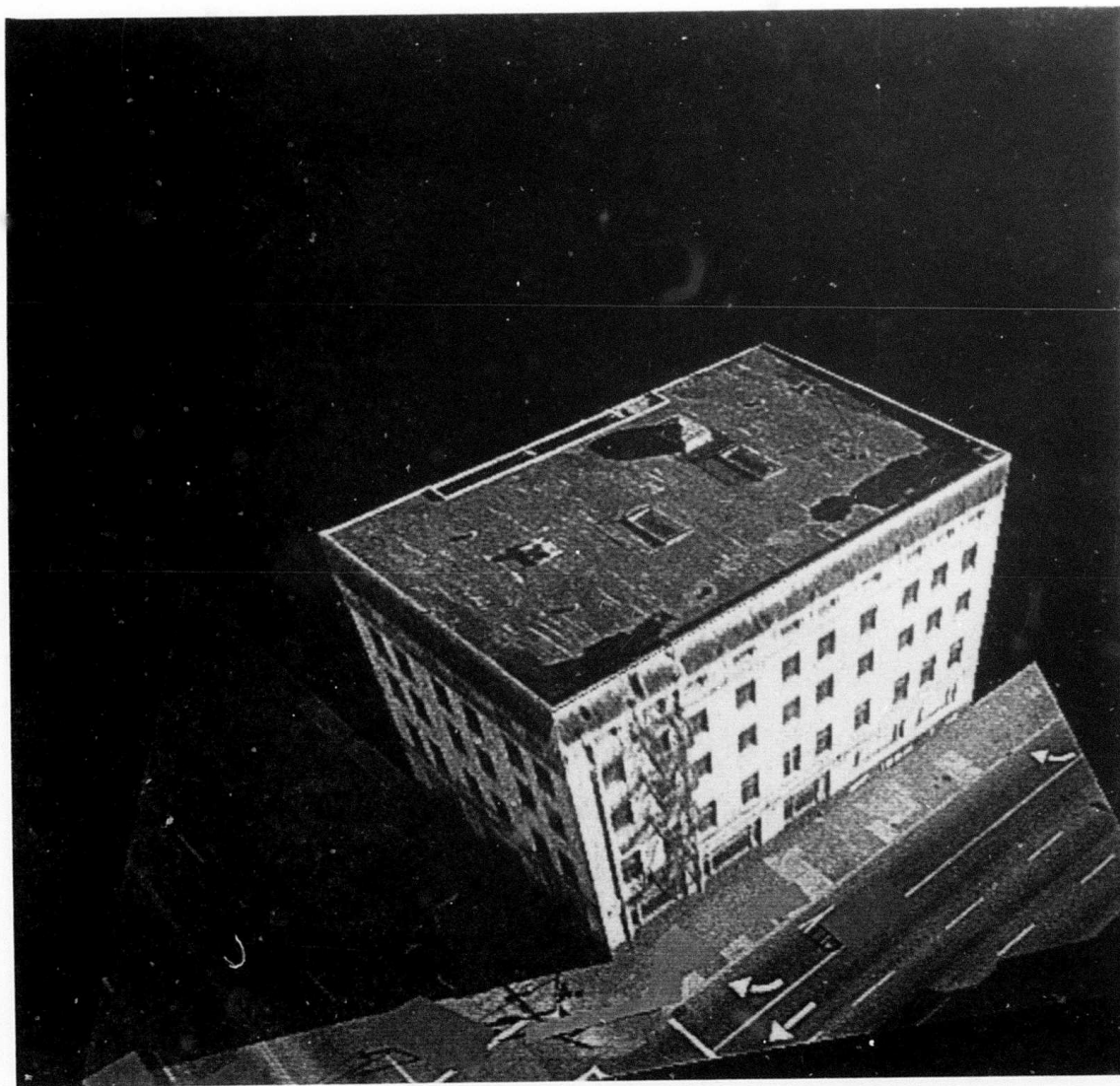Figure 6-9.   Diagonal View, 70 Foot Altitude, Moderate Tilt

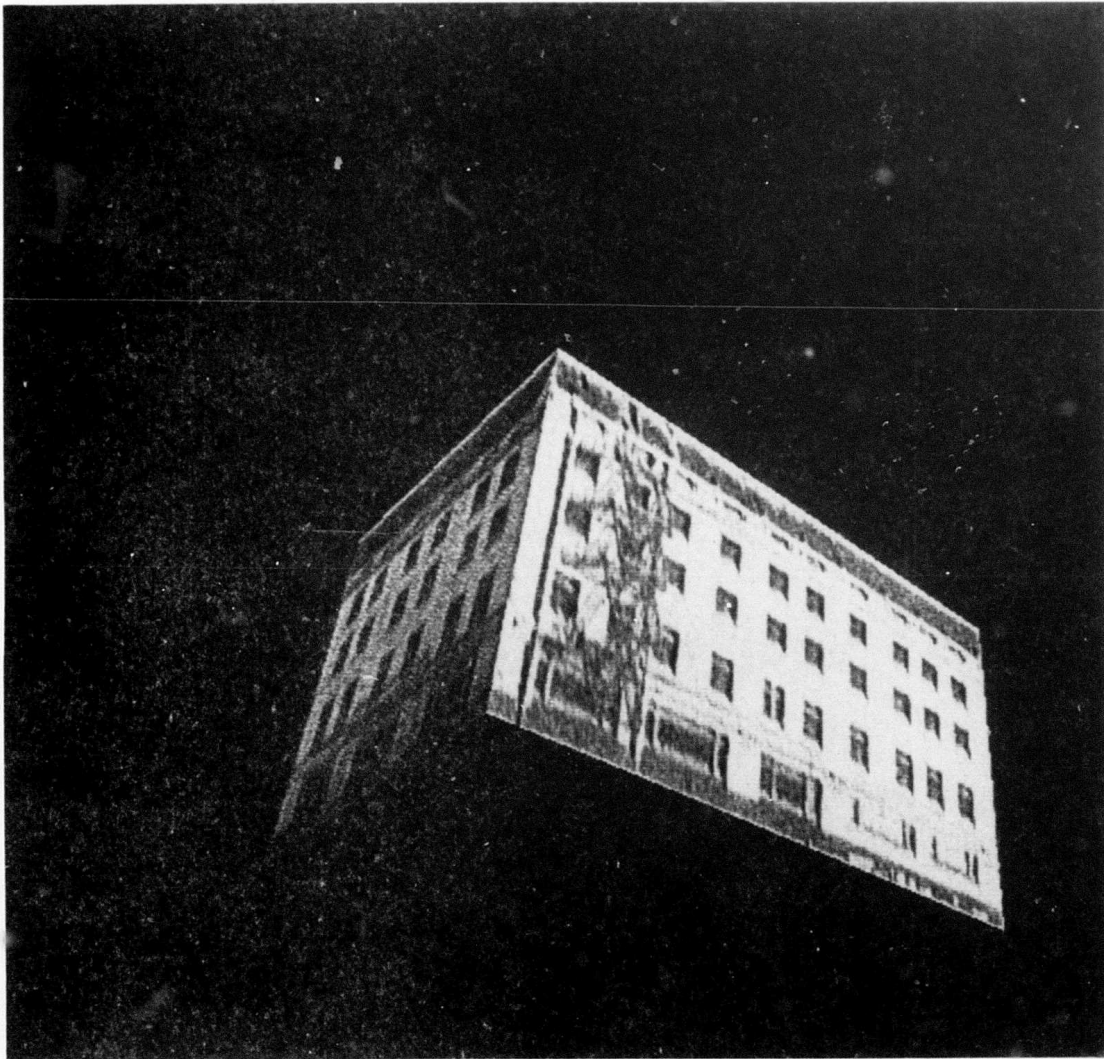Figure 6-10.   Diagonal View, 170 Foot Altitude, Significant Tilt

Figure 6-11.   Diagonal View, Below Street Level, Upward Tilt
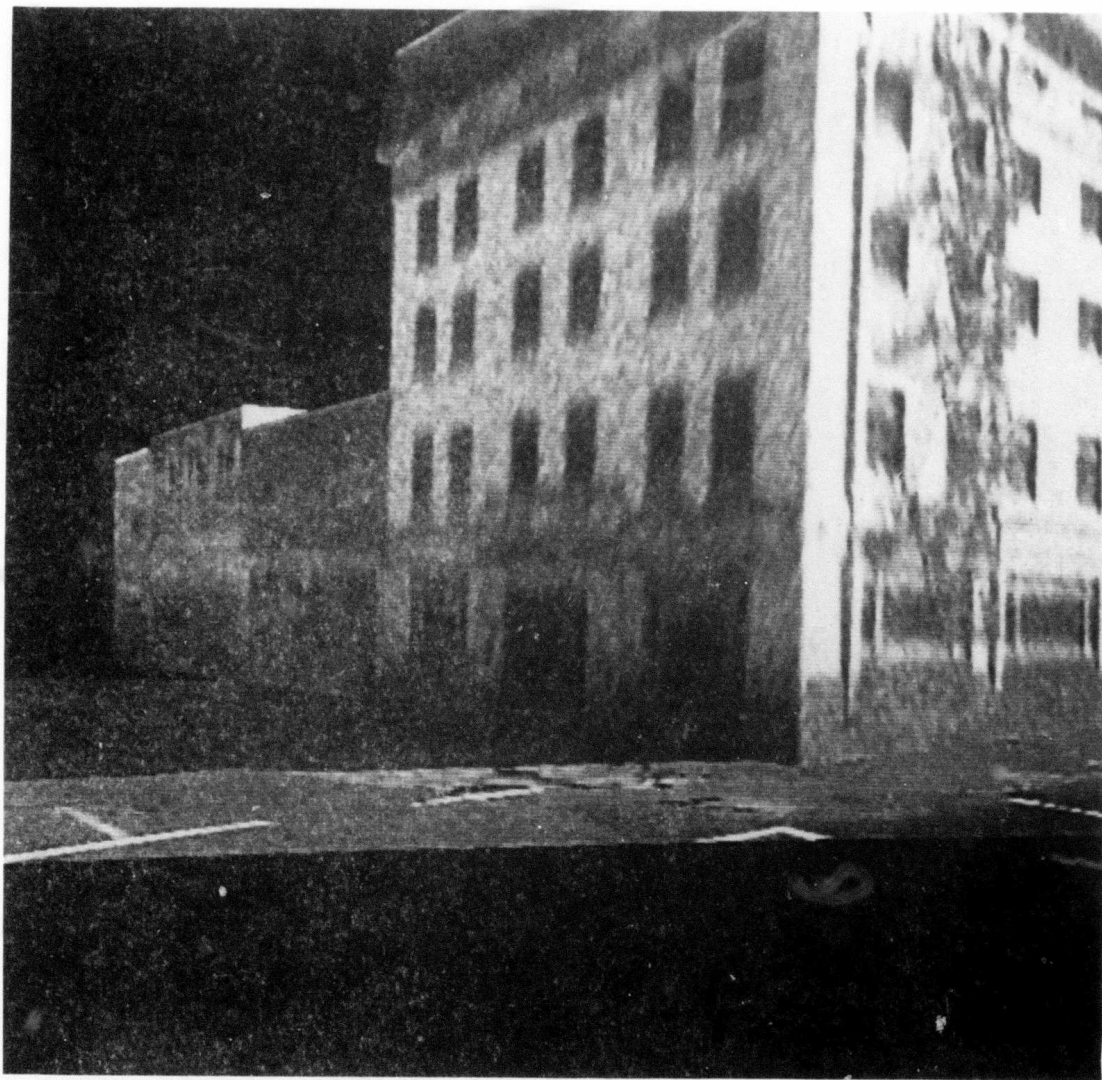
Figure 6-12. First View of Sequence

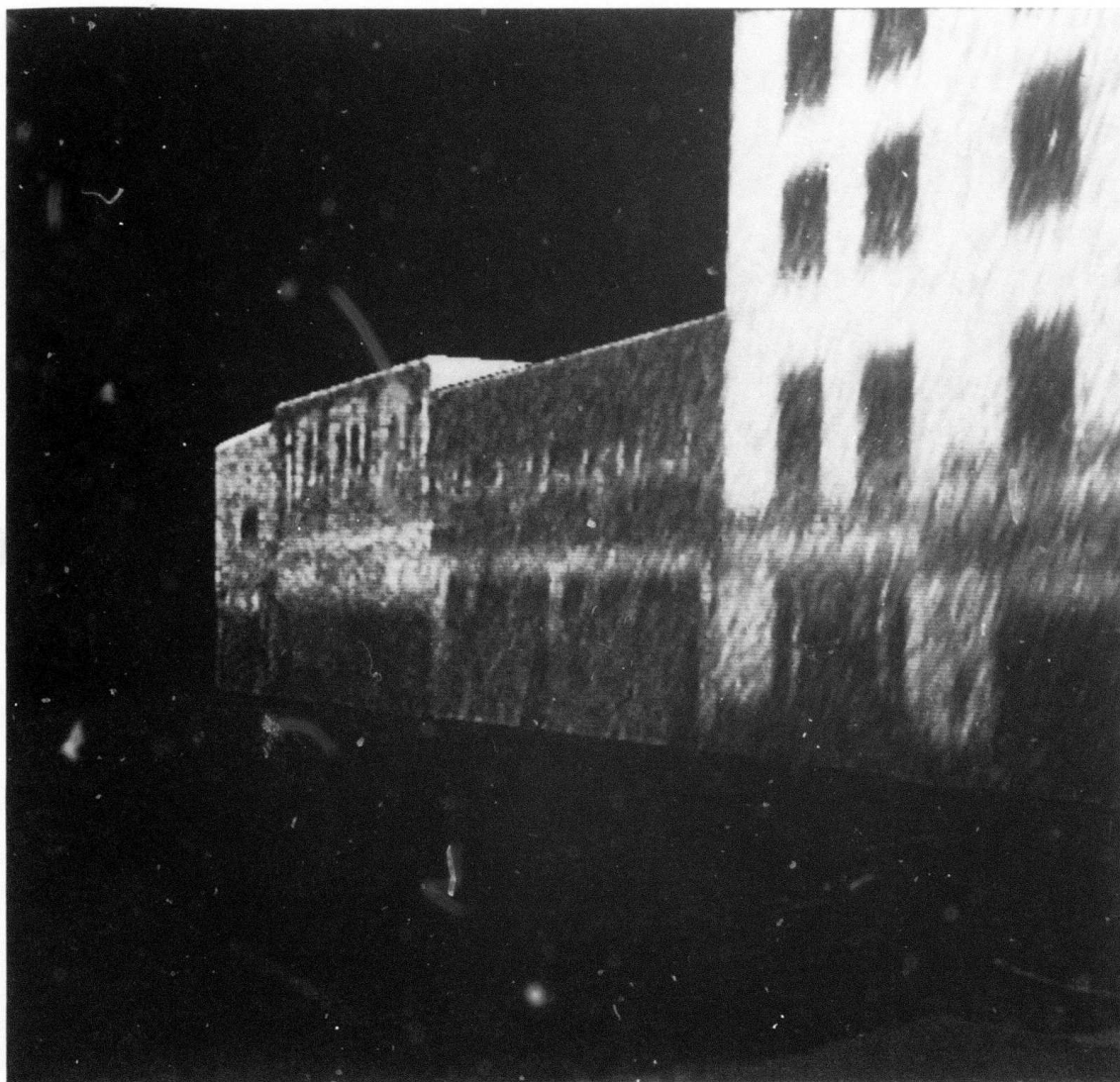Figure 6-13. Second View of Sequence

Figure 6-14. Third View of Sequence

7.   REFERENCES.

1. M. Thompson, ed., *Manual of Photogrammetry*, Vol. I, Chap. 2 (Falls Church, VA: American Society of Photogrammetry, 1966).

2. F. Moffitt, *Photogrammetry*, (International Textbook Co., 1967) pp. 260-263.

3. W. Boge, "Resection Using Iterative Least Squares," *Photogrammetric Engineering*, Vol. 31, pp. 701-714, (1965).

4. K. Rampal, "A Closed Form Solution for Space Resection," *Photogrammetric Engineering*, Vol, 45, pp. 1255-1261, (1979).

5. M. Morse, "Projective Methods," *Photogrammetric Engineering*, Vol. 32, pp. 849-855, (1966).

6. M. Aoki, M. Levine, "Computer Generation of Realistic Pictures," *Computers and Graphics*, Vol. 3, pp. 149-161, (1978).

7. J. Blinn, M. Newell, "Texture and Reflection in Computer Generated Images," *Communications ACM*, Vol. 19, No. 10, (1976).

8. I. Sutherland, R. Sproul, R. Schumacker, "A Characterization of Ten Hidden-Surface Algorithms", *ACM Computing Surveys*, Vol. 6, pp. 1-55, (1974).

8. GLOSSARY.

CAMERA MODEL. The pan (θ), tilt (ϕ), roll (ζ), focal length (f), and position (x,y,z) of a camera.

CROSS PRODUCT. The cross product of two vectors $(a_1, a_2, a_3)^T$ and $(b_1, b_2, b_3)^T$ is the vector $(a_2b_3 - a_3b_2, a_3b_1 - a_1b_3, a_1b_2 - a_2b_1)^T$.

DIGITAL IMAGE. An array of numbers that represent the intensity of an image at discrete locations over a two-dimensional grid.

GRAY LEVEL. The intensity value associated with a picture element (pixel).

IMAGE COORDINATES. The line and sample coordinate system that can be used to identify a picture element.

IMAGE MODEL. The camera model and scanning parameters associated with a digital image.

INNER PRODUCT. The inner or dot product of two vectors $(a_1, b_1, c_1)^T$ and $(a_2, b_2, c_2)^T$ is the real number $a_1a_2 + b_1b_2 + c_1c_2$.

MICROMETER (μm). One-millionth of a meter.

PERSPECTIVE TRANSFORMATION. A geometric projection of a scene formed by passing a ray from each scene point to a center of projection at a finite distance from the scene, and finding the image formed by the intersection of the rays with a plane of projection.

PICTURE COORDINATES. A coordinate system for measuring points on a picture. The coordinates are in linear units, such as feet or meters, as opposed to the line, sample system of image coordinates.

8.＿＿＿＿＿ -- Continued.

PIXEL, OR PICTURE ELEMENT. A single element of a digital image that can be identified by a (line, sample) coordinate and a gray level value.

SCANNING PARAMETERS. The scanning parameters relate image coordinates to picture coordinates. They include the linear unit measures of the sample spacing, and the x- and y- offsets between the pixel at (line 1, sample 1) and the picture coordinate origin.

SPATIAL COORDINATES. A three-dimensional coordinate system whose components are measured in linear units.